FINAL TECHNICAL REPORT

Multiscale random field-based probabilistic regional liquefaction mapping

Qiushi Chen, Charng H. Juang and Chaofeng Wang

Glenn Department of Civil Engineering, Clemson University, Clemson, SC 29634

Contact Information (Qiushi Chen)

Phone: (864) 656-3330

Fax: (864) 656-2670

Email: qiushi@clemson.edu

Website: https://cecas.clemson.edu/geomechanics/

U.S. Geological Survey

Earthquake Hazards Program

Award Number: **G17AP00044**

Project period: January 2017 – May 2018

This project is supported by the U.S. Geological Survey under Grant No. G17AP00044. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey

ABSTRACT

This one-year project aims to improve predictive models for regional liquefaction hazard mapping. In particular, the project team developed and implemented a multiscale random field model-based framework for regional liquefaction mapping. In this framework, a classical liquefaction model is adopted to assess liquefaction potentials at the site-specific scale using cone penetration test data. The statistical distribution and spatial variability of relevant soil parameters and liquefaction hazard measures are characterized and explicitly accounted for at this scale. A unique feature of this work is that the spatial correlation is extended to multiple length scales to characterize cross-scale soil spatial variability. Moreover, existing surficial geologic data are used to constrain the large-scale spatial variations of soil parameters. Both geotechnical data and geologic data are integrated using a conditional sequential simulation algorithm. The framework is then applied to map liquefaction potential index across the Alameda county site of California. The results are validated with existing knowledge and liquefaction observations, and show improved predictions when both geotechnical and geological data are appropriately accounted for in the liquefaction hazard maps.

To verify random field-based mapping approaches, a detailed three-dimension synthetic digital soil field is artificially generated along with benchmark liquefaction hazard fields. The digital field is used as a basis for assessing and verifying various random field-based models for liquefaction mapping. Different virtual field investigation plans are designed to assess the dependency of data inference and model performance upon the level of availability of sampling data. Model performances are assessed using three information theory-based measures. Results show that when sampling data is sufficient, all random field-based models examined capture fairly well the benchmark liquefaction potentials in the studied field. As the size of the sampling data decreases, the accuracy of predictions decreases for all models but to different degrees; the three-dimensional random field model gives the best result in this scenario. All random field-based models examined in this paper yield a slightly more conservative prediction of liquefaction potential over the studied field.

Ongoing work has extended the developed multiscale random field-based framework to account for uncertainties in future seismic events by integrating a recently developed model to calculate the joint distribution of peak ground acceleration and the moment magnitude of an earthquake from the latest USGS National Seismic Hazard Maps.

1 INTRODUCTION

There have been continuous efforts, many led and funded by the U.S. Geological Survey (USGS), to provide increasingly accurate liquefaction hazard maps for earthquake-prone regions in the United States. Conventional approaches for regional liquefaction evaluation and mapping often rely on surficial geology and the resulting liquefaction susceptibility or potential maps predict constant liquefaction hazard within a surficial geologic unit (Youd and Perkins, 1978; Knudsen et al., 2000; Witter et al., 2006). Subsequently, geotechnical data-based site-specific liquefaction models are integrated with geostatistical models to map liquefaction hazards (Holzer et al., 2006; Liu and Chen 2006; Baise et al. 2006; Lenz and Baise 2007; Heidari and Andrus 2010; Vivek and Raychowdhury 2014; Chen et al. 2016a&b). A more recent work in this area is to develop liquefaction models for global-scale applications (Zhu et al. 2015 & 2017). One major challenge with the incorporation of geotechnical and geology data, however, is that the surficial geology data and the geotechnical borehole data represent spatial variability of soil properties at two distinct scales, i.e., the regional scale (~km) and the site-specific scale (~m). A consistent and effective approach to account for such multiscale nature of soil spatial variability is critically important and is needed for more accurate regional liquefaction modeling.

This one-year project aims to improve predictive models for regional liquefaction hazard assessment. In particular, the project team develops and validates a hybrid geotechnical and geological data-based framework for multiscale regional liquefaction potential mapping (Wang and Chen, 2018; Chen et al., 2018; Wang et al., 2018). The framework builds on recent advances on multiscale random field models that explicitly account for the spatial variability of soil parameters across different length scales (Baker et al., 2011; Chen et al., 2012; Chen et al., 2016a&b; Liu et al., 2017). Both geotechnical data and surficial geology data-based liquefaction potential estimations are integrated into the mapping process through a conditional simulation algorithm. Liquefaction potential maps of the Alameda County of California are generated and validated against previously published hazard maps and liquefaction observations from the 1989 Loma Prieta earthquake. To verify and compare the effectiveness of different random field model-based mapping methods, a synthetic digital soil field is developed and a benchmark liquefaction potential field is used to quantify the performance of different mapping approaches (Juang et al., 2018; Wang et al., 2017).

The remaining of the report will detail (1) the hybrid geotechnical and geological data-based framework for multiscale liquefaction mapping, and (2) the synthetical digital field-based model validation. Main conclusions and ongoing work are discussed in the summary section.

2 GEOTECHNICAL AND GEOLOGICAL DATA-BASED MULTISCALE LIQUEFACTION HAZARD MAPPING

2.1 Overview of the methodology

A schematic illustration of the developed framework is shown in Figure 1. In this project, liquefaction hazard is quantified and mapped in terms of the liquefaction potential index (LPI) (Iwasaki et al., 1978 & 1982). The proposed methodology, however, is general and can be applied to map other measures of liquefaction hazard, such as liquefaction-induced ground settlement and lateral spread.

As shown in Figure 1, the liquefaction hazard map is generated by taking into account two

types of LPI data. The first type, termed the 'primary data', is evaluated using a geotechnical-based LPI model. In this project, the Robertson and Wride (1998) cone penetration test (CPT)-based liquefaction model is adopted. As shown in Figure 1(a), CPT data are collected within the study area upon which the primary LPI are calculated. The second type, termed the 'secondary data', is obtained based on secondary information such as a surficial geological map (Figure 1(a)). As shown in Figure 1(b), the distribution of primary LPIs within each geological unit is characterized. Then, as shown in Figure 1(c), the secondary LPI data within each geological unit are generated based on the characterized distributions of primary LPIs. The role of the secondary data is to constrain and improve the primary data-based LPI map such that the final map conforms to the large-scale geological boundaries in the analysis region.

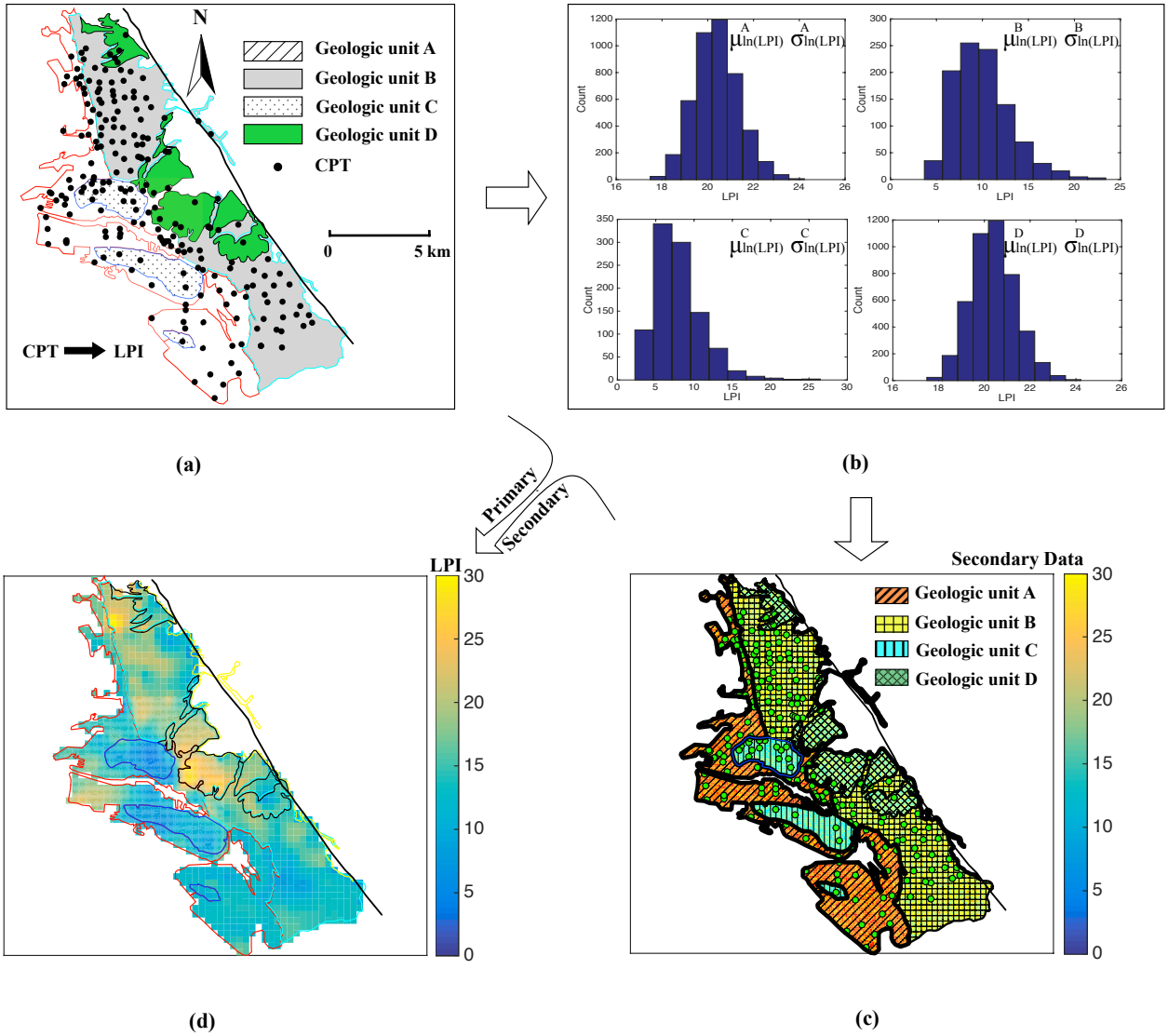


Figure 1. Regional mapping of LPI with geological constraints: (a) collect CPT data and calculate primary LPI data at CPT soundings; (b) characterize the distribution of LPIs within each geologic unit; (c) generate secondary LPI data within each geological unit based on the characterized parameters of distribution; (d) random field realization of LPI over the whole region conditioned upon both primary and secondary data.

Based on the primary CPT-based LPI data and the secondary geology-based LPI data, multiscale random field models are developed to generate realizations of LPIs across the region of interest, as shown in Figure 1(d). Coupled with Monte Carlo simulations, uncertainties associated with the generated liquefaction hazard maps can also be obtained. Various quantities of interest related to liquefaction hazard can be obtained and results will be presented and discussed in the later part of this report.

2.2 Project site and data

The site used to demonstrate the developed framework is the Alameda County site of California shown in Figure 2. The availability of engineering data and the extensive past liquefaction studies in this region make it an ideal site to test and validate the proposed methodology. Information on engineering geology of this site has been compiled by Helley and Graymer (1997) and briefly summarized in Holzer et al. (2006). This region is divided into three broad northwest-southeast trending regions, parallel to the Hayward Fault – the most important seismic source in this area. Bedrock is exposed at the surface of the northeast land. The central area, immediately southwest of the bedrock, consists of the Holocene and Pleistocene alluvial fan deposits. The area next to the central area – southwest of the original natural shoreline – is primarily underlain by the artificial fill that rests on younger San Francisco Bay mud.

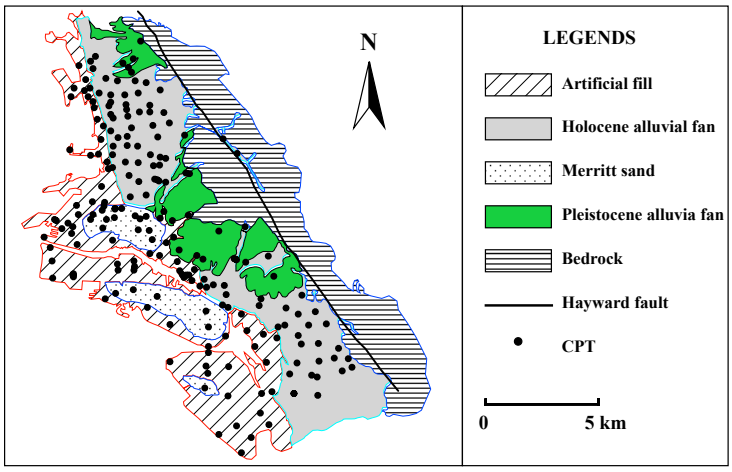


Figure 2. Map of the Alameda County site, surficial geology and locations of 210 CPT soundings (black dots). The surficial geology map is generated based on information compiled by Holzer et al. (2006).

A total of 210 CPT sounding profiles are collected from the U.S. Geological Survey (USGS) online CPT database (https://earthquake.usgs.gov/research/cpt/). Water table information is directly obtained from each CPT sounding record. For CPT soundings without water table information, simple interpolation is used. For the unit weight of soil, constant values of γ_m=15.0 kN/m^3 for moist soil above the water table and γ_{sat}=19.4 kN/m^3 for saturated soil below the water table are assumed. However, when data are available, it would be more accurate to use varying unit weights along soil layers. A hypothetical magnitude M_w= 7.1 earthquake is assumed with a constant peak horizontal acceleration at the ground surface, a_{max}= 0.5 g. This combination of a_{max} and M_w has been assumed in a previous liquefaction mapping study (Holzer et al., 2006) for a hypothetical earthquake event on the nearby Hayward Fault. The assumption of a constant a_{max} was justified on the basis that the outcrop area of the surficial geological unit generally parallels and is close to the Hayward Fault (Holzer et al., 2006). Alternatively, varying peak ground

acceleration (PGA) values obtained from the latest USGS seismic hazard map could be adopted for the estimation of liquefaction hazard.

2.3 CPT-based primary LPI data

In this work, the classical CPT-based liquefaction model by Robertson and Wride (1998) is implemented to calculate the liquefaction potential of a soil layer, where two variables –the cyclic stress ratio (CSR) and the cyclic resistance ratio (CRR) – are evaluated. Details of this classical liquefaction model are included in the Appendix of this report.

With CSR and CRR defined, the factor of safety (FS) against liquefaction at a particular depth z can be calculated as

$$\mathrm{FS} = \frac{\mathrm{CRR}}{\mathrm{CSR}} \tag{1}$$

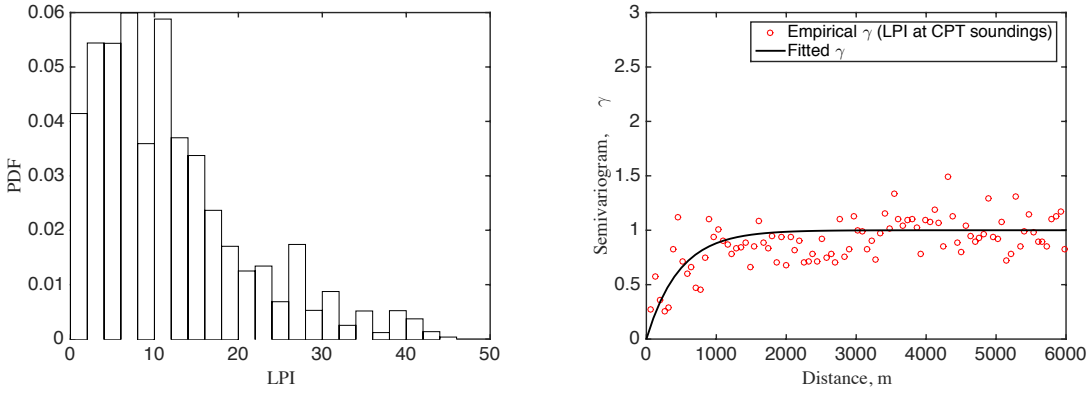
To quantify liquefaction hazard at a particular location, the factor of safety is integrated over the top 20 m of soil to obtain an averaged index termed the liquefaction potential index (LPI) (Iwasaki et al., 1978, 1982). Details of the LPI calculation are included in the Appendix of this report. The LPI will be used as the primary variable to be mapped over the project region through multiscale random field models.

For the project site and given a hypothetical earthquake event (M_w = 7.1 and a_{max} = 0.5g), the primary LPI values are calculated at 210 CPT locations. The histogram of the calculated 210 LPI values is plotted in Figure 3(a). To assess the spatial correlations of the primary LPI data, the empirical semivariogram $\hat{\gamma}(\mathbf{h})$ is calculated as (Goovaerts, 1997)

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \left[z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h}) \right]^2 \tag{2}$$

where $N(\mathbf{h})$ is the number of pairs of data z located a vector $\mathbf{h}$ apart (i.e., a lag bin $\mathbf{h}$). In the actual computation, a small tolerance (e.g., 10 – 20% of the distance $\mathbf{h}$) is usually added to lag bins to accommodate unevenly spaced sample points. Also, it is often convenient to use a scalar distance measure h for the calculation of semivariogram (e.g., Wang et al., 2017). Figure 3(b) plots the calculated empirical semivariogram based on LPIs at 210 CPT locations. Given the sample semivariogram, a weighted least square method (Cressie, 1985) is implemented to fit an analytical semivariogram model, shown as the solid line in the plot. In this work, an exponential model is fitted, given as

$$\gamma(h) = 1 - e^{(-h/470)} \tag{3}$$



(a) (b)

Figure 3. Primary LPI data at 210 CPT locations in the Alameda County: (a) histogram in probability density function (PDF); (b) semivariograms.

2.4 Geology-based secondary LPI data

Knowledge of surficial geology is an important piece of information for regional liquefaction mapping as it typically provides broader area coverage and information on large-scale material heterogeneity. Potentially, any surficial geology-dependent liquefaction information (e.g. previous liquefaction observations, regional geology-based liquefaction hazard map) could be used to derive secondary LPI data that essentially enforce geological constraint to the generated liquefaction hazard map. In this section, a simplified procedure is presented to utilize the calculated primary LPI values and the surficial geological map to derive secondary LPI data. The method consists of the following steps:

Step 1. Identify the boundaries of each geological unit within the study region.

Step 2. Group the primary LPI values by geological units and characterize their statistical distributions.

Step 3. Generate random variable realizations of secondary LPI values within each geological unit according to the characterized or assumed statistical distributions. Once generated, the secondary data will be kept constant for the following random field realizations.

Step 4. Assign the generated secondary LPI values to a predefined grid, which will be integrated into the conditional sequential simulation algorithm as secondary data.

For the project site, the surficial geology map of the Alameda County has been shown in Figure 2, from which the boundaries of the geologic units are identified. As summarized above, the next step is to group primary LPI values by geologic units and characterize their statistical distribution. Figure 4 shows the histograms of primary LPIs at 210 CPT soundings grouped by the four main geologic units in the study region, i.e., the artificial fill, the Pleistocene fan, the Holocene fan and the Merritt sand. It can be seen from Figure 4 that the distributions of LPIs within the artificial fill, the Holocene fan and the Merritt sand geological units can be fitted with a lognormal distribution, whereas the primary data are too sparse to conclude a basic distribution for the Pleistocene fan unit. In the following, a lognormal distribution will be assumed when generating secondary LPI data from a distribution within each surficial geologic unit.

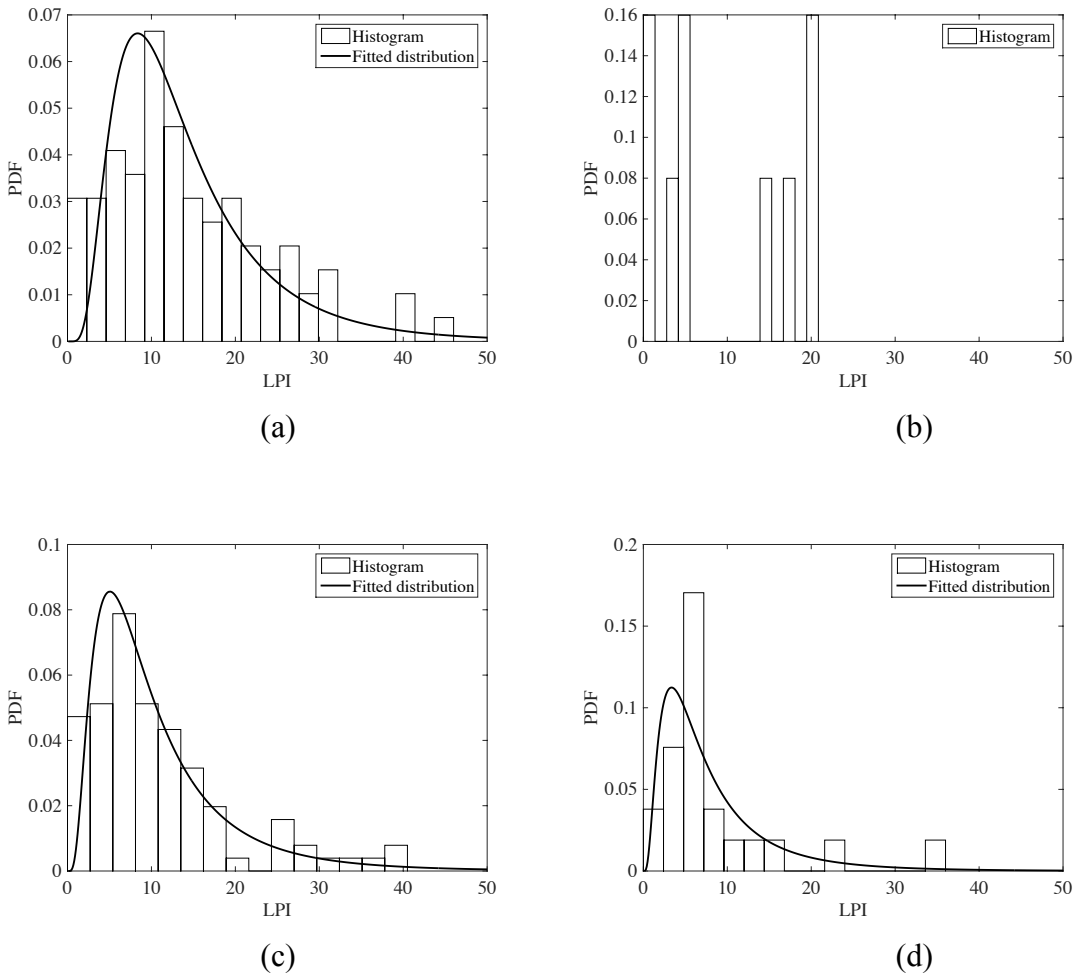


Figure 4. Histograms of primary LPIs for each geologic unit: (a) artificial fill; (b) Pleistocene fan; (c) Holocene fan; (d) Merritt sand.

In addition to histograms, box plots and cumulative frequency curves are also used to understand the primary LPI values grouped by geological unit and are shown in Fig. 5. It can be seen from the box plots (Figure 5(a)) that the median LPI values for each geological unit are ordered from high to low in this manner: artificial fill, Holocene fan, Merritt sand, Pleistocene fan. The cumulative frequency plots (Figure 5(b)) show that, for a given LPI value, the artificial fill unit has the highest cumulative frequency values, while the Merritt sand unit has the lowest frequency values. The curve for the Pleistocene fan unit shows some abrupt changes due to a very limited amount of data points. Previously, such cumulative frequency curves have been used to assign a constant probability of liquefaction value to the corresponding geological unit (Holzer et al., 2006).

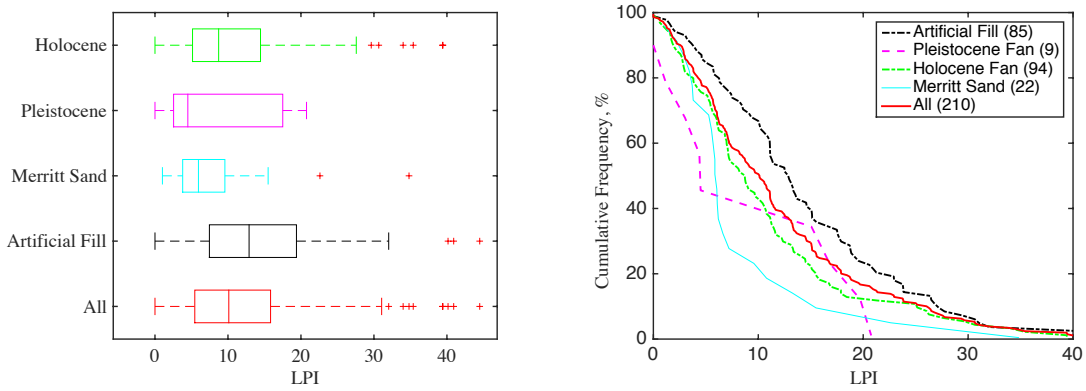


Figure 5. LPI data by geologic units: (a) box plots and (b) cumulative frequency of primary LPI data for each geologic unit. The number in the bracket is the number of CPT soundings in that geologic unit.

Basic statistical parameters of the primary LPIs for each geological unit are summarized in Table 1. Consistent with the above analysis, artificial fill has the highest mean value of LPI, followed by the Holocene fan. The parameters for Pleistocene fan and Merritt sand are very close to each other.

Table 1. Statistics of primary LPIs by surficial geologic units.

Geologic unit	Mean, μ	Standard deviation, σ
Artificial fill	14.58	9.47
Pleistocene fan	9.48	8.43
Holocene fan	10.87	8.87
Merritt sand	8.36	7.65

Once the distribution type and statistical parameters are determined, secondary LPI values for each geological unit are obtained from random field realizations and are assigned to a predefined grid. The results are shown in Figure 6 for all geological units in the study region. Such secondary LPI values will be incorporated into the multiscale random field models described in the next section. It is worth noting that the above analysis groups geotechnical data based on surficial geological units. Geological units, however, are three-dimensional features. If the subsurface information is available, it may be more appropriate to group geotechnical data according to three-dimensional features.

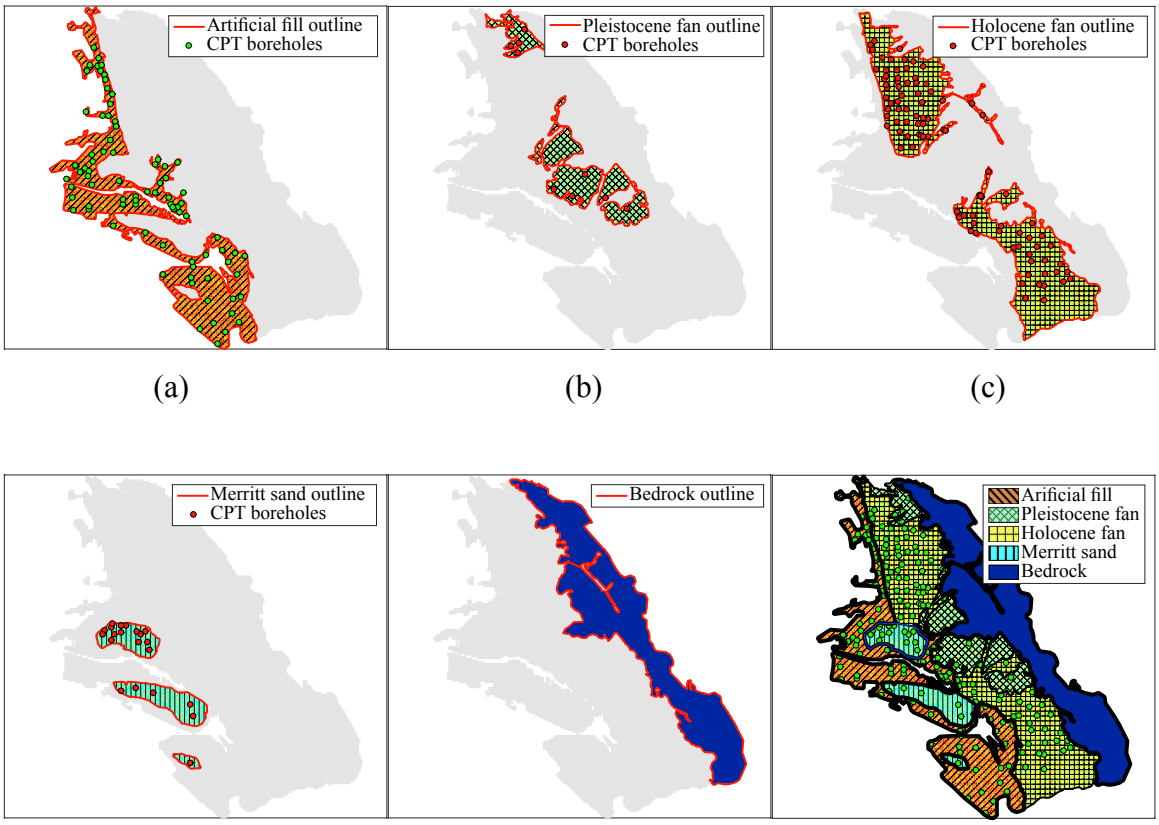


Figure 6. Generation of secondary LPI data for each surficial geologic unit: (a) articial fill; (b) Pleistocene fan; (c) Holocene fan; (d) Merritt sand; (e) bedrock; (f) all

2.5 Multiscale random field models

In this section, novel multiscale random field models are presented to integrate both the primary and secondary LPI data and to map liquefaction hazards (in LPIs) across the study site. Key components of the multiscale random field models are described in this section.

2.5.1 Conditional sequential simulation algorithm

A conditional sequential simulation algorithm is implemented in this project to generate random field realizations of LPIs across the project site. This algorithm integrates and preserves multiple sources of known data (e.g. primary and secondary LPI data). In this algorithm, the realization of a random variable Z_n is represented by a joint distribution as follows

$$\begin{bmatrix} Z_n \\ \mathbf{Z}_p \\ \mathbf{Z}_s \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_n \\ \boldsymbol{\mu}_n \\ \boldsymbol{\mu}_n \end{bmatrix}, \begin{bmatrix} \boldsymbol{\sigma}_n^2 & \Sigma_{np} & \Sigma_{ns} \\ \Sigma_{pn} & \Sigma_{pp} & \Sigma_{ps} \\ \Sigma_{sn} & \Sigma_{sp} & \Sigma_{ss} \end{bmatrix} \right) \tag{4}$$

where $\sim N(\boldsymbol{\mu},\Sigma)$ denotes the vector of random variables following a joint normal distribution with the mean vector μ and the covariance matrix Σ; Z_n is the random variable to be generated with the expected value μ_n; Z_p is the vector of previously generated or known primary random

variables with the vector of expected values μ_p; Z_s is a vector of secondary random variables with the vector of expected values μ_S; σ_n is the standard deviation of Z_n; Σ is the covariance matrix with subscripts 'n', 'p' and 's' denoting 'next', 'previous primary' and 'secondary', respectively. The individual terms in the covariance matrix is defined as

$$\mathrm{COV}[Z_i, Z_j] = \boldsymbol{\rho}_{Z_i,Z_j} \boldsymbol{\sigma}_{Z_i} \boldsymbol{\sigma}_{Z_j} \tag{5}$$

where $\boldsymbol{\rho}_{Z_i,Z_j}$ is the correlation between two elements Z_i and Z_j within the random field at any scale with a standard deviation of $\boldsymbol{\sigma}_{Z_i}$ and $\boldsymbol{\sigma}_{Z_j}$, respectively.

Given the joint distribution in Eq. (4), the distribution of the random variable Z_n, conditional upon all previously simulated and known primary and secondary data, is given by a univariate normal distribution with the updated mean and variance as

$$(Z_n \,|\, (Z_p, Z_s)) \sim N(\tilde{\boldsymbol{\mu}}_n, \tilde{\boldsymbol{\sigma}}_n) \tag{6}$$

with

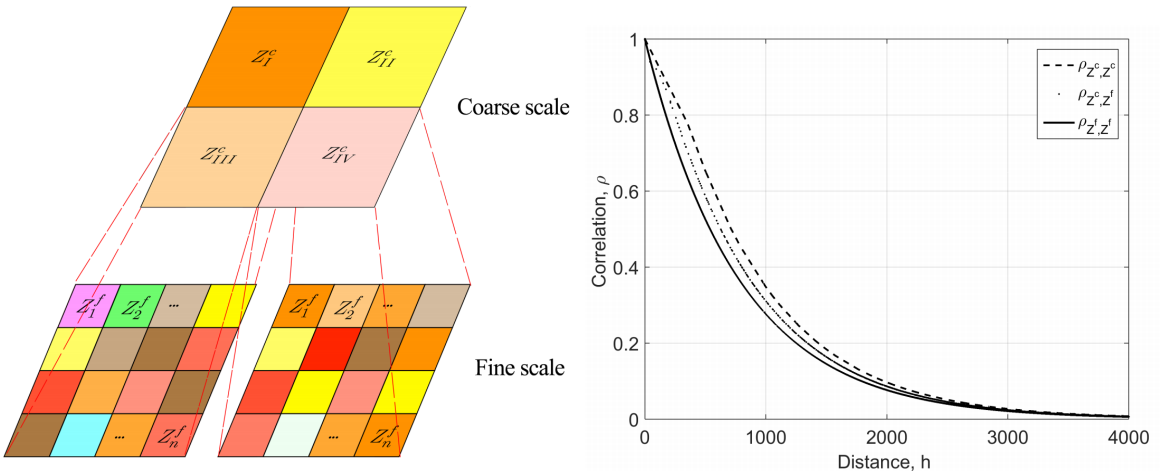
$$\tilde{\boldsymbol{\mu}}_n = \begin{bmatrix} \Sigma_{np} & \Sigma_{ns} \end{bmatrix} \begin{bmatrix} \Sigma_{pp} & \Sigma_{ps} \\ \Sigma_{sp} & \Sigma_{ss} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Z}_p \\ \mathbf{Z}_s \end{bmatrix} \tag{7}$$

$$\tilde{\boldsymbol{\sigma}}_n = \boldsymbol{\sigma}_n^2 - \begin{bmatrix} \Sigma_{np} & \Sigma_{ns} \end{bmatrix} \begin{bmatrix} \Sigma_{pp} & \Sigma_{ps} \\ \Sigma_{sp} & \Sigma_{ss} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_{pn} \\ \Sigma_{sn} \end{bmatrix} \tag{8}$$

where the symbols in Eq. (7) and (8) have been defined after Eq. (4). The value of a random variable Z_n at an unsampled location is drawn from the above joint distribution. Once generated, Z_n becomes a data point in the vector Z_p to be conditioned upon by all subsequent simulations. This process is repeated by following a random path to each unknown location until all the values in the field have been simulated – that is, a map of the primary variable for the region of interest is generated.

2.5.2 Characterization of spatial correlations across scales

In-situ test data such as the CPT data used in this project provide detailed information about soil properties at specific locations. Due to the spatial dependence, higher confidence on the generated soil properties near sampling locations is expected, which leads to higher confidence in the predicted liquefaction susceptibility. In other words, it is plausible, adjacent to a sampling location, to generate a higher resolution random field to account for the soil spatial variability at smaller scales. In this study, the conventional procedure to characterize soil spatial variability is extended to incorporate multiple scales of resolution. The main challenge entails developing a consistent representation of fine-scale and coarse-scale random fields while maintaining appropriate spatial correlation structures across scales.



(a) Multiscale elements (b) Multiscale correlations

Figure 7. Illustration of cross-scale relations and the resulting correlation function (ρ).

In this work, two scales of interest are considered, i.e., Scale *c* for a coarse scale and Scale *f* for a fine scale, as shown in Figure 7(a). The numbering for coarse elements is in Roman numerals, while for fine elements Arabic numbers. The coarse scale is defined as the average of all fine-scale points within its area (element)

$$Z_I^c = \frac{1}{N}\sum_{i=1,i\in I}^{N} Z_i^f \tag{9}$$

where N = number of fine scale points within a coarse-scale area (element) *I*. All variables described in the previous section are considered to exist at the fine scale. Using such a relation, the expression for the variances and spatial correlations of coarse-scale and cross-scale material properties can be explicitly derived

$$\boldsymbol{\rho}_{Z_{c(a)},Z_{c(b)}} = \frac{\Sigma_{i=1}^{N}\Sigma_{k=1}^{N}\boldsymbol{\rho}_{Z_{fi(a)},Z_{fk(b)}}}{\sqrt{\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\boldsymbol{\rho}_{Z_{fi(a)},Z_{fj(a)}}}\sqrt{\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\boldsymbol{\rho}_{Z_{fi(b)},Z_{fj(b)}}}} \tag{10}$$

$$\boldsymbol{\rho}_{Z_f,Z_{c(a)}} = \frac{\Sigma_{i=1}^{N}\boldsymbol{\rho}_{Z_f,Z_{fi(a)}}}{\sqrt{\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\boldsymbol{\rho}_{Z_{fi(a)},Z_{fj(a)}}}} \tag{11}$$

where $\rho_{Z_{c(a)},Z_{c(b)}}$ is the correlation between two coarse scale elements '*a*' and '*b*'; $\rho_{Z_f,Z_{c(a)}}$ is the correlation between a fine scale element and coarse element '*a*'. The derived spatial correlation curves across different scales are plotted in Figure 7(b).

Given the multiscale spatial dependence specified by Eqs. (10) and (11) and an inferred or assumed probability distribution of the random variable, the conditional sequential simulation approach Eqs (4-8) is employed to generate random field realizations of variables of interest.

2.5.3 Covariances across data types

To perform the conditional sequential simulation in Eq. (6), three covariances must be determined: one for the primary variable, one for the secondary variable and a cross-covariance

describing the relationship between these variables. With relatively sufficient amount of primary LPI data calculated from CPT soundings, the covariance of the primary variable can be easily obtained from the inferred spatial correlation between primary data. However, direct calculation of the secondary and cross-covariance can be challenging. In this work, one simplified approach is adopted based on the Markov–Bayes hypothesis described by Goovaerts (1997) to derive the secondary and cross-covariance by calibrating them to the primary covariance as (Goovaerts, 1997; Moysey et al., 2003)

$$\mathrm{COV}_s(\mathbf{h}) = \begin{cases} |B| \cdot \mathrm{COV}_p(\mathbf{h}) & for\ \mathbf{h} = 0 \\ B^2 \cdot \mathrm{COV}_p(\mathbf{h}) & for\ \mathbf{h} > 0 \end{cases} \tag{12}$$

$$\mathrm{COV}_{ps}(\mathbf{h}) = B \bullet \mathrm{COV}_p(\mathbf{h}) \tag{13}$$

where B is the Markov–Bayes coefficient; $\mathrm{COV_p}$ is the covariance for the primary variable, $\mathrm{COV_s}$ is the covariance for the secondary variable and $\mathrm{COV_{ps}}$ is the covariance between the primary and the secondary variables; h is the distance vector separating two random variables. The Markov–Bayes coefficient B generally varies between 0 and 1 when primary and secondary variables are positively correlated. Its value affects the relative importance of primary data and secondary data and this effect will be illustrated in the results section.

In Wang and Chen (2018), a simple procedure to calibrate the coefficient B is presented, which is based on a calibration procedure recommended by Deutsch et al. (1998) such that the value B is determined as the difference between two conditional expectations as follows

$$B = E_1 - E_0 \tag{14}$$

where the two conditional expectations are defined as

$$E_1 = E[P(Z_s \le Z) \mid Z_p \le z] \tag{15}$$

$$E_0 = E[P(Z_s \le z) \mid Z_p > z] \tag{16}$$

where E is the expectation operator; E is the expectation; Z_s is the secondary variable (e.g. the geological data-based LPI value) and $P(Z_s \leq Z)$ is the probability of Z_s less than or equal to a threshold value z (e.g. a given LPI threshold value); Z_p is the primary variable (e.g. the geotechnical data-based LPI value). The conditional expectation E_1 will be close to 1 if the primary and secondary data support each other – that is, the two data predict similar liquefaction hazard levels. The conditional expectation E_0 will be close to 1 if the primary and secondary data contradict each other – that is, the two data predictions of liquefaction hazard contradict each other.

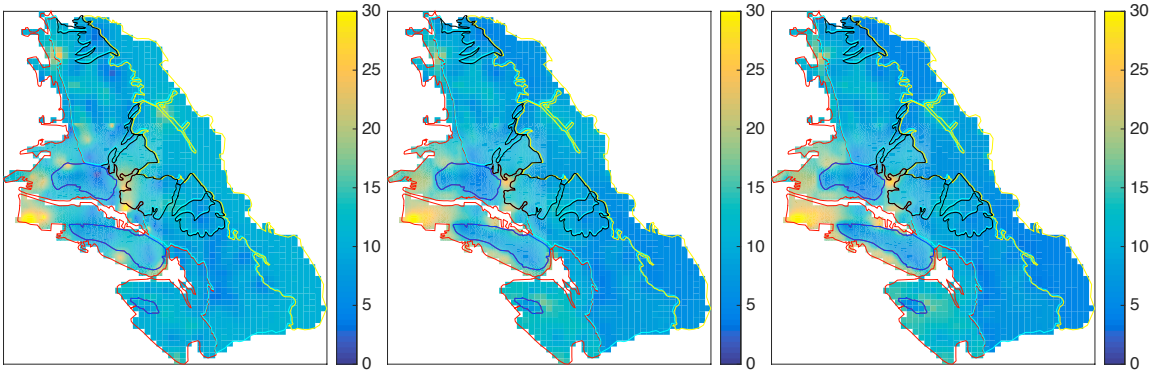
The effectiveness of calibration using Eq. (14), however, depends on the availability of secondary data. In general, more accurate secondary data would yield better calibration results. If there are not enough secondary data, one can choose the B parameter based on the confidence of the secondary data. A higher B value means the secondary data has a higher weight on the final generated maps.

2.6 Results and discussions

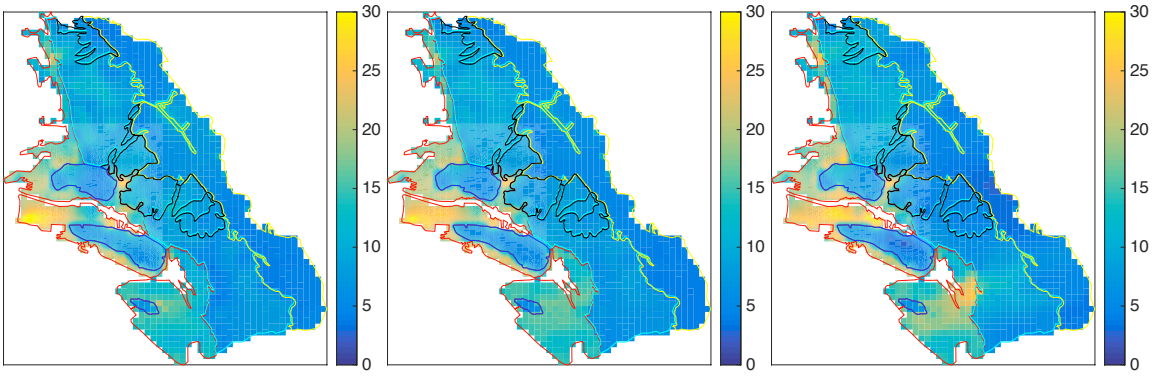
2.6.1 Monte Carlo analysis

For regional liquefaction hazard mapping, the multiscale random field models are coupled with Monte Carlo simulations to obtain the expected liquefaction hazard across the region and to perform probabilistic analysis on quantities of interest. To investigate the influence of the Markov–Bayes coefficient B introduced in Eq. (14), six B values are used – namely, 0, 0.1, 0.4, 0.5, 0.73 and 0.9. For each B value, 1000 Monte Carlo simulations are performed. A hypothetical earthquake event (M_w = 7.1, a_{max} = 0.5 g) is assumed for all simulations.

Figure 8 shows maps of expected LPI values for all six cases of Markov–Bayes coefficient B. Each of the six maps is obtained by averaging results from 1000 Monte Carlo simulations. As shown in Eqs. (12) and (13), the Markov–Bayes coefficient B is essentially a 'scaling' factor between the primary covariance and the secondary covariance matrices. The larger the coefficient B, the stronger influence the secondary data has on the generated LPI maps. In this work, the secondary LPI data come from geological information. Therefore, as the value of B increases, the geological boundaries become more distinguishable in the resulting LPI maps, as evidenced in Figure 8. When no secondary data are incorporated – that is, B = 0 – no geological constraint is applied to the LPI map. Such a case is shown in Figure 8(a) and the resulting LPI map might be inaccurate. For instance, lacking geological constraint, the bedrock unit (refer to Figure 2 for the geological unit map of the study region) is predicted to have LPI values around 10, which is usually classified as a high liquefaction severity class (refer to Table 4). This contradicts the common knowledge that the bedrock is not prone to liquefaction. By applying secondary LPI data, this type of incorrect classification could be mitigated.



(a) (b) (c)



(d) (e) (f)

Figure 8. Maps of expected LPI values for all six cases of Markov-Bayes coefficient B. B*=0 is the case without secondary LPI data. B*=0.73 is the case where the coefficient B is calibrated. Each map is obtained by averaging results from 1000 Monte Carlo simulations: (a) B* = 0; (b) B=0.1; (c) B=0.4; (d) B=0.5; (e) B*=0.73; (f) B=0.9

The case of B = 0.73 is obtained following the calibration procedure described in Eq.(14) based on an LPI threshold of 5, which is the value separating the moderate and the high liquefaction severity class (refer to Table 4). It should be noted that the selection of LPI threshold affects the calibrated B value. For the Alameda County site, the threshold value of 5 is chosen as it appropriately separates liquefaction-prone and non-liquefied units. As shown in Figure 8(e), the LPI map with the calibrated B = 0.73 manifests reasonable spatial variations of LPI that comply with the known data while preserving boundaries of different geological units. It should be noted that the calibrated B value can be used as a starting point for further adjustments. For instance, if there is a relatively high confidence in the quality of secondary data, or if the secondary data are known to have greater impact on liquefaction, the coefficient B could be increased. If there is very little knowledge or low confidence about the secondary data, a lower value of B could be adopted.

The results obtained and shown in Figure 8 can be utilized to quantify liquefaction hazard and to calculate various quantities of interest. As an example, following the procedure proposed by Holzer et al. (2006), cumulative frequency (CF) distributions of LPIs are calculated to assess the liquefaction potentials of each geological unit. The CF distributions obtained with different B values are plotted in Figure 9. For a given geological unit, Holzer et al. (2006) related the percentage area predicted to undergo liquefaction during a given earthquake shaking scenario to the value of CF corresponding to LPI = 5. shows that the artificial fill geological unit has the highest CF values for a given LPI and is therefore most likely to liquefy. The Merritt sand unit has the lowest CF values for a given LPI and is predicted to be the most resistant to liquefaction. The impact of coefficient B on the CF distributions is relatively small.

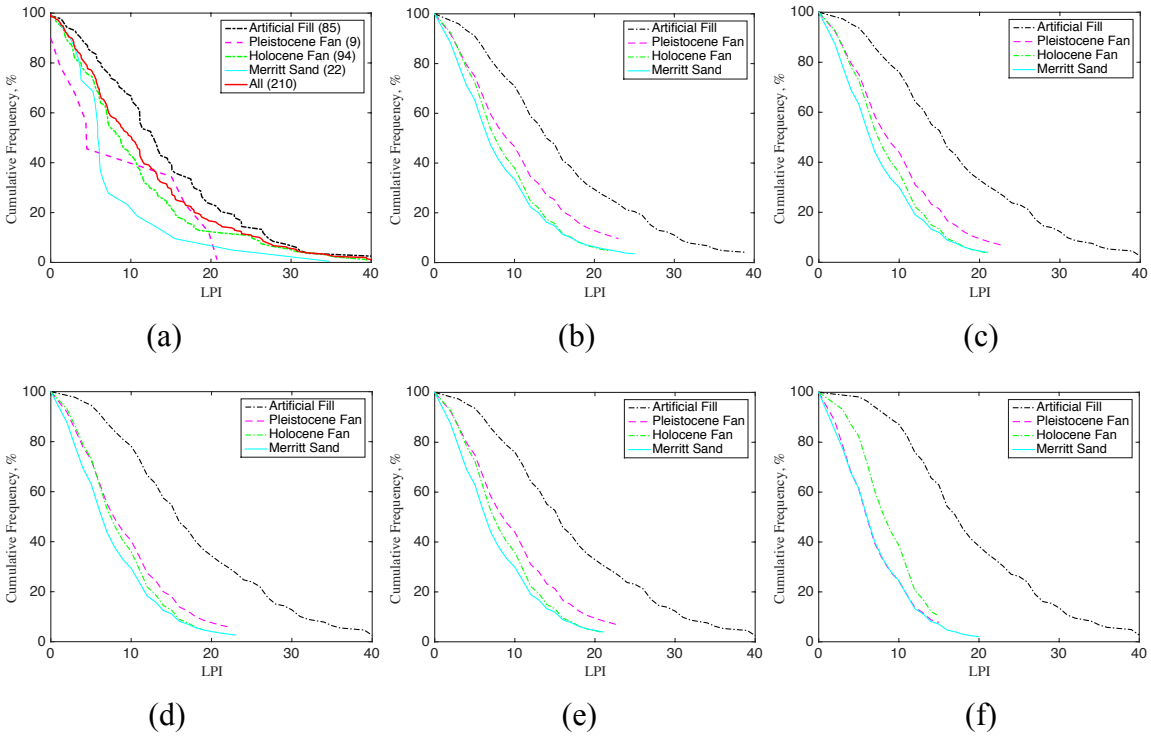
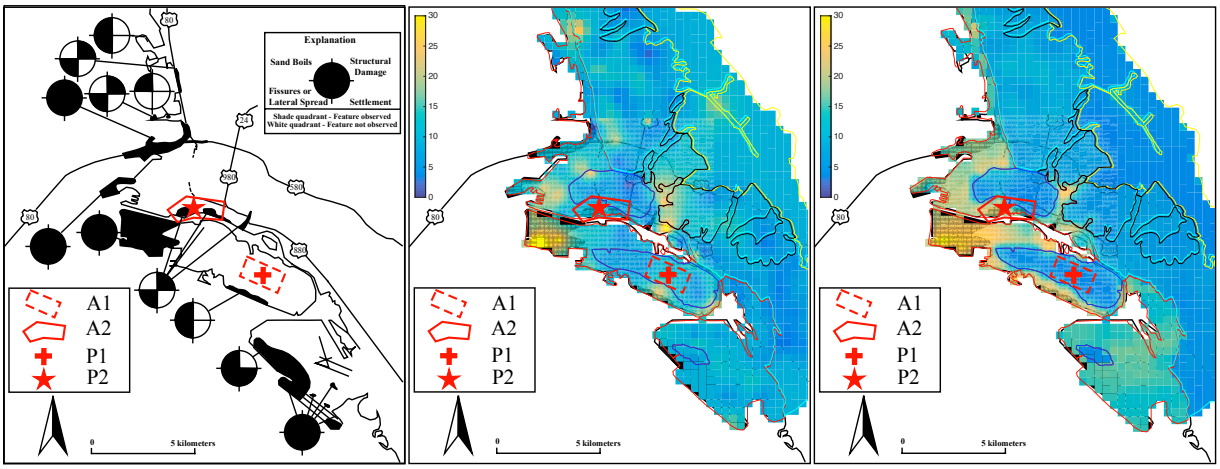


Figure 9. Cumulative frequency distributions of LPI: (a) known data; (b) B = 0.1; (c) B = 0.4; (d) B = 0.5; (e) B = 0.73; (f) B = 0.9

2.6.2 Validation with prior knowledge and observations

As a modest validation of the framework, maps of the expected liquefaction potential hazards are compared with observations of liquefaction following the 1989 Loma Prieta earthquake (Kayen et al., 1998). As shown in Figure 10, most of the observed liquefied areas (along the coastline and in the artificial fill unit) are also predicted to have high LPI values by the proposed framework. As previously mentioned, the LPI map with the calibrated Markov–Bayes coefficient B = 0.73 predicts the bedrock unit as non-liquefiable (LPI ≈ 0), which is consistent with the fact that no liquefaction phenomenon was observed during past earthquakes.



(a) (b) (c)

Figure 10. Comparison of liquefaction observation from the 1989 Loma Prieta earthquake with LPI maps generated by the proposed framework: (a) liquefaction observations (modified from Kayen et al., 1998); (b) LPI map with B = 0; (c) LPI map with B = 0.73

3 VERIFICATION OF RANDOM FIELD-BASED MAPPING APPROACHES USING A SYNTHETIC DIGITAL SOIL FIELD

3.1 Overview

While geostatistical tools and random field models are increasingly used in liquefaction mapping studies, a systematic assessment and verification of different approaches to account for spatial variation and dependence of soil properties or liquefaction potentials are missing and the implications of various random field-based mapping approaches are unknown. The main challenge is the lack of sufficient data, and therefore lack of knowledge about the soil properties and liquefaction potentials of the field. Moreover, in situ test data are typically sampled at selected and sometimes clustered locations, resulting in additional complexities to assess random field-based model performance.

To overcome these challenges, in this section, an extremely detailed three-dimensional synthetic digital soil field is artificially generated and used as a basis to assess and verify various random field-based approaches for liquefaction mapping. Soil properties of interest (e.g., the CPT tip resistance) are known at every location in the synthetic field. The benchmark liquefaction potential fields can, therefore, be obtained for any given hypothetical earthquake event. Moreover, different virtual field test plans are designed to assess their effects on data inference and model performances.

Given such an extensive amount of information, this study will assess and verify various common and uncommon random field-based liquefaction mapping approaches. In particular, this study will assess: (1) the performance and effectiveness of various approaches in mapping quantities of interest (e.g., soil properties, LPIs) over studied region; (2) the effect of amount of field data on the relative performances of different approaches, and (3) the optimal random field-based liquefaction model for mapping liquefaction hazards. This study aims to provide insights on approaches that are commonly used to account spatial variability and dependence in random field-based liquefaction mapping studies.

3.2 Three random field-based approaches for liquefaction mapping

In this section, three random field-based approaches for liquefaction mapping are described along with details of the random field model used in the following simulations. Again, the classical CPT-based liquefaction model proposed by Robertson and Wride (1998) and subsequently updated by Robertson (2009) is used to evaluate liquefaction potential of a soil layer. The liquefaction hazard is then quantified and mapped over a region in terms of the liquefaction potential index (LPI) (Iwasaki et al., 1978&1982). Details of the CPT-based liquefaction model and the LPI calculations are included in the Appendix section. Depending on how the spatial dependence and variation are integrated in the mapping process, three common and uncommon random field-based approaches will be assessed and verified: the averaged index approach, the two-dimensional (2D) local soil property approach and three-dimensional (3D)

local soil property approach, which are denoted herein as M1, M2 and M3, respectively. A schematic detailing these three approaches is shown in Figure 11.

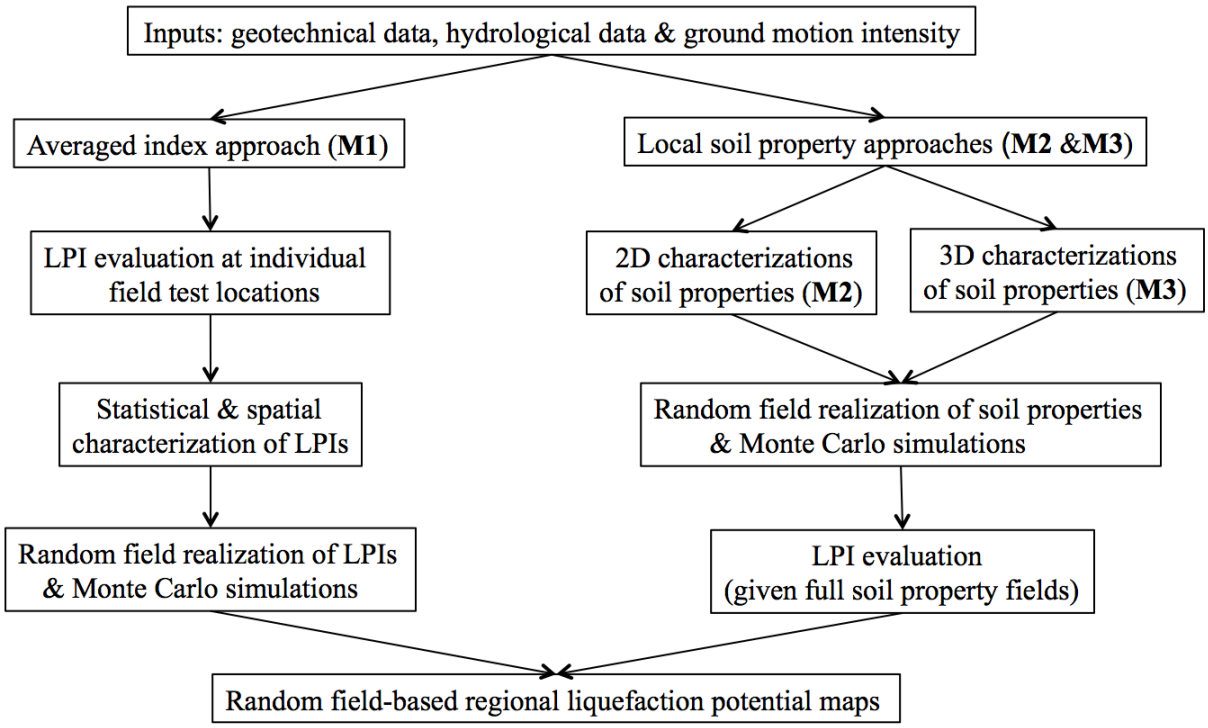


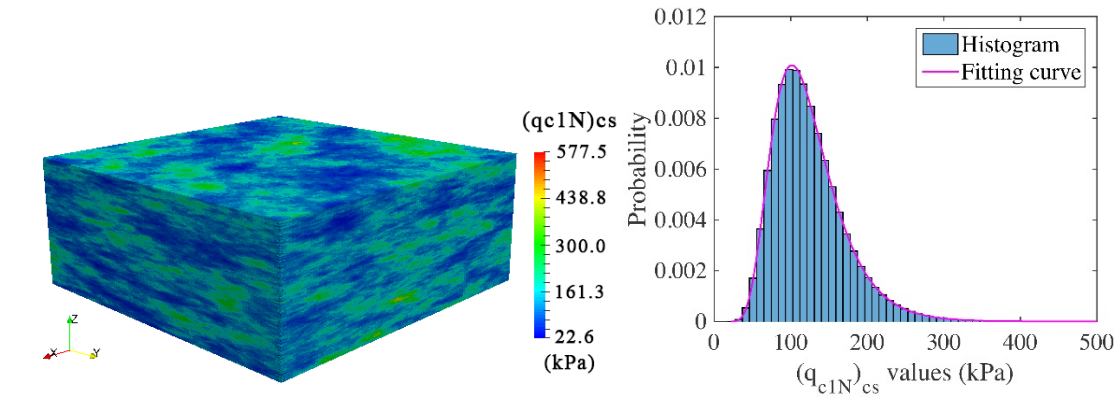
Figure 11. Three random field-based liquefaction mapping approaches.

As shown in the figure, the averaged index approach (M1) treats the LPI as the spatially correlated random variable while the 2D and 3D local soil property approaches (M2 and M3) treat the soil properties of interest (e.g., tip resistance from CPT test) as spatially correlated random variables. In the 2D local soil property approach, the random field of soil properties is generated layer-by-layer considering the horizontal correlation within each layer. In contrast, the 3D local soil property approach considers both horizontal and vertical spatial correlations. All approaches will rely on the random field models and Monte Carlo simulations to generate regional liquefaction potential maps.

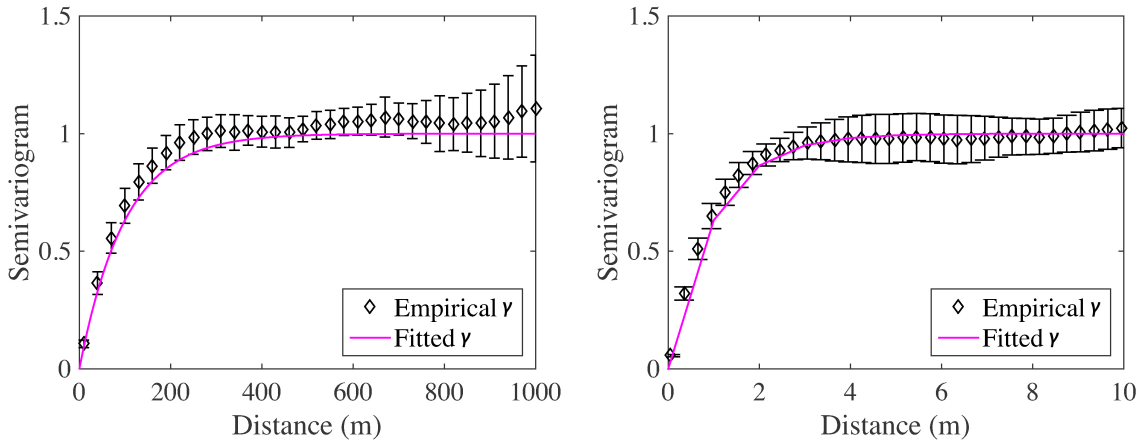
3.3 Synthetic digital soil field and benchmark LPI field

The dimension of the synthetic digital soil field is set as 1000 ×1000 ×20 m (width × length × deep) and a soil element size is correspondingly set as 10 ×10 ×0.05 m. There are a total of 4,010,000 soil elements in the field. The depth of the digital field (20 m) corresponds to the integration depth in LPI calculation detailed in the Appendix section. The soil element is assigned to have a thickness of 0.05 m to match the vertical sampling interval of a typical CPT test.

Within this field, a three-dimensional and spatially correlated clean sand equivalent tip resistance $(qc1N)_{cs}$ field is generated and its values are assigned to each soil element as shown in Figure 12. The parameters used to generate the synthetic field are based on the experience gained through the spatial analysis of the CPT database in Alameda County of California (Chen et al., 2016b).



(a) Three-dimensional view (b) Histogram



(c) Semivariogram in XY plane (d) Semivariogram in YZ plane

Figure 12. The three-dimensional view, the histogram and semivariograms of the synthetic digital (qc1N)cs field. The empirical semivariograms (c) and (d) show both the mean values as well as the error bars (± standard deviation) from the averaging of all layers.

The $(qc1N)_{cs}$ of the digital soil field is assumed to follow a lognormal distribution, and the spatial correlation of the field is specified as isotropic in the horizontal plane and anisotropic in the vertical plane. The histogram of the $(qc1N)_{cs}$ is shown in Figure 12(b),with the mean μ and the variance σ^2 as 123.98 kPa and 2182.68 kPa, respectively. The semivariogram $\gamma(h)$ in the XY plane and YZ plane are respectively shown in Figure 12(c) and Figure 12(d). The error bars (± one standard deviation s) represent the variance of empirical semivariogram in the 401 XY planes and 100 YZ planes. The magenta line is the fitted g by Eq. (6), and the correlation length $a_x = a_y = 82.59$ m, $a_z = 0.915$ m. For simplicity, the synthetic digital $(qc1N)_{cs}$ field is denoted as the "true" $(qc1N)_{cs}$ field for use in subsequent model verifications. It should be noted that the true distribution and spatial structure of this digital soil field are unknown to random field-based liquefaction modeling and mapping, same as in the case of a real soil field. The lognormal and assumptions made on spatial correlation are for the convenience of generating the digital field.

To calculate the benchmark liquefaction potential index (LPI) field, the following input data for liquefaction model and for a hypothetical earthquake scenario are used: the moist unit weight of the soil γ_m is taken as a constant at 15 KN/m^3, the saturated unit weight γ_{sat} is 19 KN/m^3, the groundwater table GWT is at 3 m below ground surface, the maximum horizontal acceleration at the ground surface a_{max} = 0.3g and the moment magnitude M_w = 7.0.

The resulting benchmark LPI field is shown in Figure 13. The benchmark "true" LPI field for the hypothetical earthquake scenario (a_{max} = 0.3g and M_w=7.0). It will be used as the benchmark liquefaction potential field for further verification and is denoted as the "true" LPI field. According to the severity class of liquefaction listed in Table 4 by Sonmez (2003), most areas of the field are classified as "high" (IV) or "very high" (V) under the hypothetical earthquake scenario.

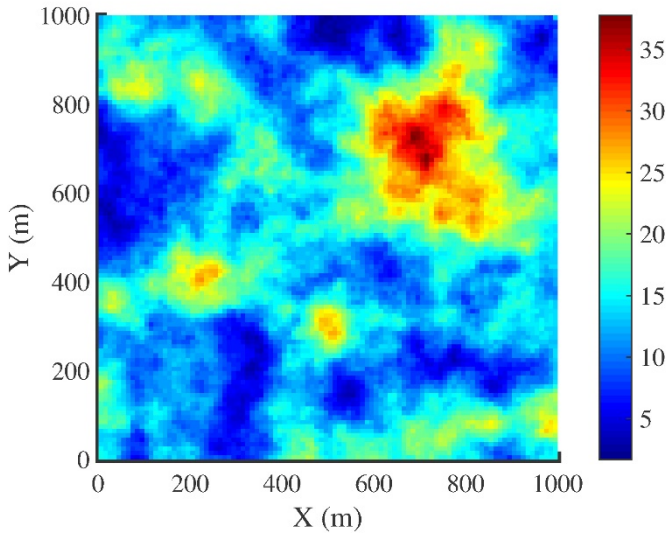


Figure 13. The benchmark "true" LPI field for the hypothetical earthquake scenario (a_{max} = 0.3g and M_w=7.0)

3.4 Procedure for model verification

3.4.1 Virtual site investigation plans

As suggested by Webster and Oliver (1992), a sample size of 100 should give acceptable confidence to estimate semivariograms of soil properties. Two investigation plans are designed in this paper to compare the model performances under the scenarios of sufficient and insufficient sample size. As shown in Figure 14, plan #1 is designed with a total of 225 evenly - spaced CPT soundings, where the $(qc1N)_{cs}$ is extracted from the digital soil field at each sounding location. As a comparison, plan #2 has only 36 evenly-spaced CPT soundings and is used to gauge the random field model performance under insufficient test samples.

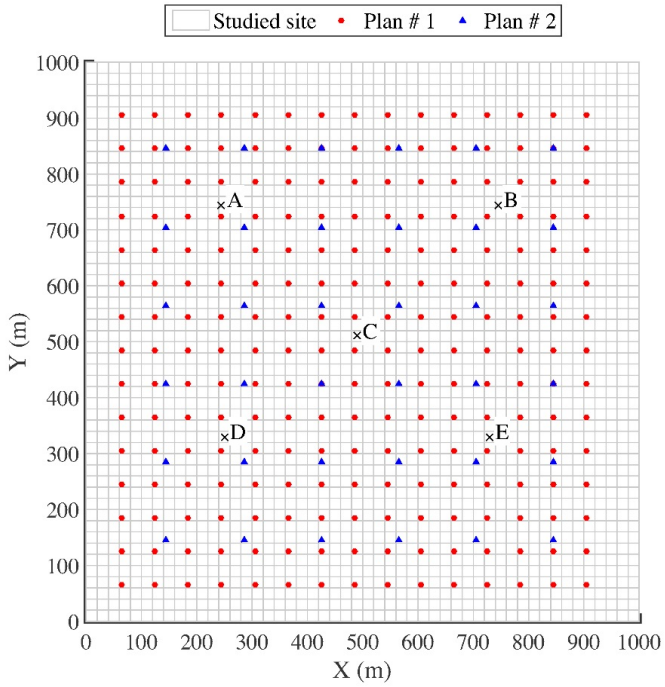


Figure 14. Layouts of the virtual site investigation plans (locations A, B, C, D and E are marked for site-specific comparisons).

3.4.2 Methods to assess model performance

The model performances of random field-based liquefaction models are assessed using the "true" data from the synthetic digital soil field and the benchmark liquefaction potential field (true fields). The models are assessed for two aspects: 1) histogram assessments to check if the random field models can simulate the data distribution of the true fields, and 2) semivariogram assessments to verify if the random field models can capture the spatial variability of the true fields. For $(qc1N)_{cs}$ field, the comparisons of $(qc1N)_{cs}$ profiles at specific locations are made to verify the random field-based liquefaction model performances. And for LPI field, the cumulative frequency plot and differences between true and simulated fields are assessed to evaluate the model performance.

In addition, three information theory-based measures are adopted to quantitively assess model performances, i.e., the mean absolute percentage error (MAPE), the root mean square deviation (RMSD) and the bias factor (Armstrong and Collopy, 1992; Prasomphan and Mase, 2013; Kung et al., 2007; Juang et al., 2012).

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(X_{\mathrm{true}})_i - (X_{\mathrm{sim}})_i}{(X_{\mathrm{true}})_i}\right| \tag{17}$$

$$\mathrm{RMSD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[(X_{\mathrm{true}})_i - (X_{\mathrm{sim}})_i]^2} \tag{18}$$

$$\text{Bias factor} = \frac{1}{n}\sum_{i=1}^{n}\frac{(X_{\text{sim}})_i}{(X_{\text{true}})_i} \tag{19}$$

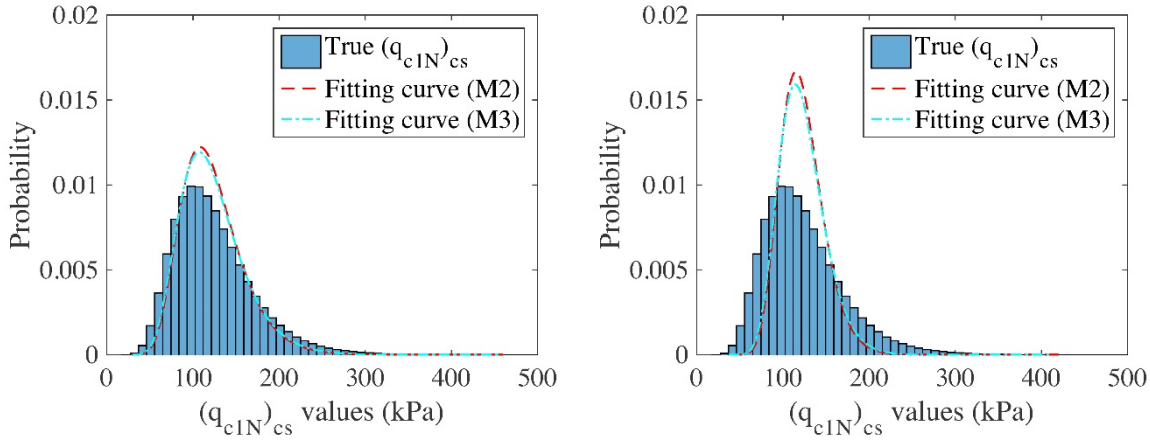
where n is the number of data; i is the i^{th} data; X is the model output value, e.g., the LPI or $(qc1N)_{\text{cs}}$ value in this paper; X_{true} is the true LPI or $(qc1N)_{\text{cs}}$ value, and X_{sim} is the simulated or predicted value. Smaller MAPE or RMSD value indicates a better model performance. For the bias factor, a value of greater than 1 means the model overestimates the true field, a value of less than 1 means an underestimation, and a value of 1 means an unbiased prediction.

3.5 Verification results and discussions

Following the procedure in the previous section, results of random field-based liquefaction models by the averaged index approach, and the 2D and 3D local soil property approaches (M1, M2, and M3) are assessed and verified in this section. Unless otherwise stated, results of the random field-based liquefaction models are the averaged values based on 1000 Monte Carlo simulations.

3.5.1 Assessment and verification - soil property fields

The histograms of the true and simulated $(qc1N)_{\text{cs}}$ values for both investigation plans are plotted in Figure 15. The blue bins represent the true $(qc1N)_{\text{cs}}$ histogram, and the red dash lines and cyan dash-dot lines represent histogram fitting curves for simulated $(qc1N)_{\text{cs}}$ values using M2 and M3, respectively. It can be seen from Figure 15(a) that both random field models predict the statistical distribution of the true soil property field well, providing that sufficient field data (investigation plan #1) are available to infer model parameters. On the other hand, the model performance deteriorates for the case with insufficient field data (investigation plan #2), as shown in Figure 15(b). The differences between predictions using 2D (M2) and 3D (M3) local soil property approaches are almost negligible.

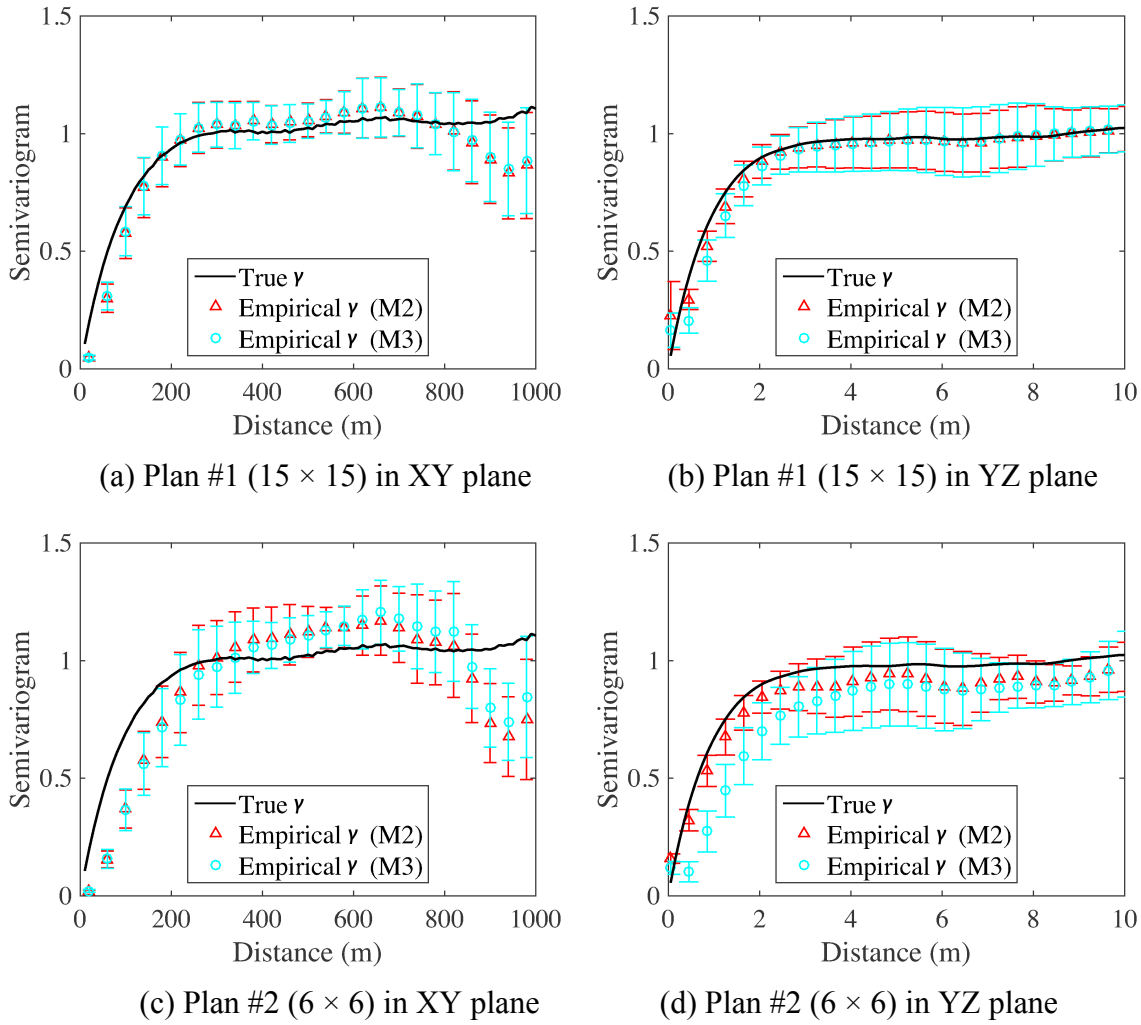


(a) Plan #1 (15 × 15) (b) Plan #2 (6 × 6)

Figure 15. Histograms of true and simulated $(q_{c1N})_{\text{cs}}$ fields of M2 and M3 for both investigation plans

The ability of random field models to capture the underlying spatial structure of the soil property field is also examined. Empirical semivariograms of the true $(qc1N)_{\text{cs}}$ field and the simulated $(qc1N)_{\text{cs}}$ fields are shown in Figure 16. The red triangles and cyan circles represent the mean values of the calculated empirical semivariograms by M2 and M3 in the XY plane and in

the YZ plane, respectively. The error bars indicate ± one standard deviation from the mean. It can be seen from the plots that for investigation plan #1, both models capture the spatial structure of the soil property field well. For investigation plan #2, the semivariograms of M2 and M3 deviate from the trend of the true semivariogram, which is not surprising as insufficient data yield a less accurate estimate of model parameters.



(a) Plan #1 (15 × 15) in XY plane

(b) Plan #1 (15 × 15) in YZ plane

(c) Plan #2 (6 × 6) in XY plane

(d) Plan #2 (6 × 6) in YZ plane

Figure 16. Semivariograms of true $(qc1N)_{cs}$ field and simulated $(qc1N)_{cs}$ fields of M2 and M3 for the both investigation plans.

To quantitatively assess model performances, the three measures introduced in Eqs. (17), (18) and (19), i.e., the MAPE, RMSD and the bias factor, are calculated and summarized in XX for the simulated $(qc1N)_{cs}$ fields by the local soil property approaches (M2 and M3). Note that in the averaged index approach (M1), $(qc1N)_{cs}$ field is not needed and therefore, no result from M1 is presented in XX. Smaller MAPE and RMSD values mean better performance, and bias factor closer to one means more accurate model. MAPE and RMSD values for both local soil property approaches (M2 and M3) are relatively small compared with the mean (123.98 kPa) and variance (2182.68) of the true field, which indicates a relatively good prediction. M3 performs slightly better than M2. For the bias factors, all model results yield slightly greater than one bias factor,

which means the random field-based models slightly overpredict. The sampling size also affects the prediction accuracy, simulations with sufficient field data (plan #1) yield better results. For all cases considered, the 3D local soil property approach (M3) outperforms the 2D local soil property approach (M2).

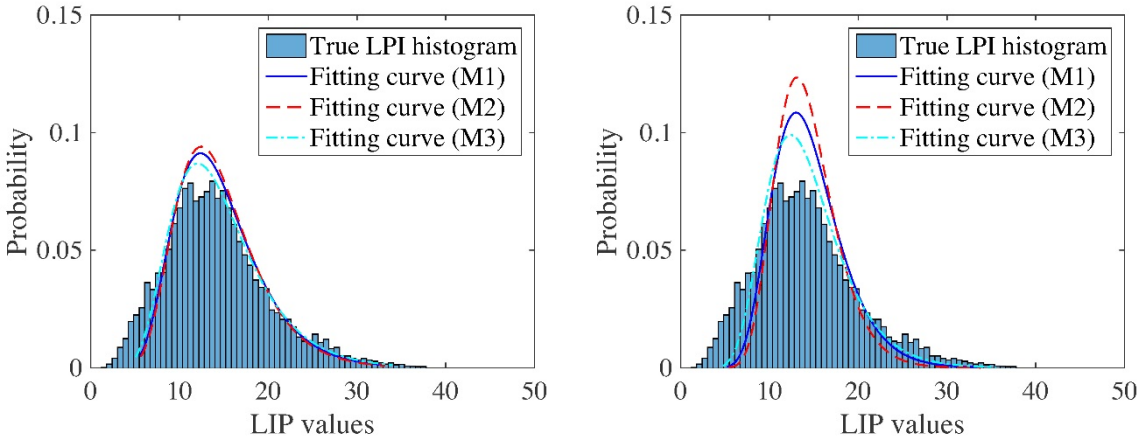
Table 2. The criteria index for the $(qc1N)_{cs}$ random fields

Index	Approach 1 (M1)		Approach 2 (M2)		Approach 3 (M3)	
	Plan #1	Plan #2	Plan #1	Plan #2	Plan #1	Plan #2
MAPE	NA	NA	0.147	0.238	0.136	0.226
RMSD	NA	NA	24.569	36.480	22.887	35.157
Bias factor	NA	NA	1.029	1.071	1.024	1.066

Note: NA means not available

3.5.2 Assessment and verification: liquefaction potential fields

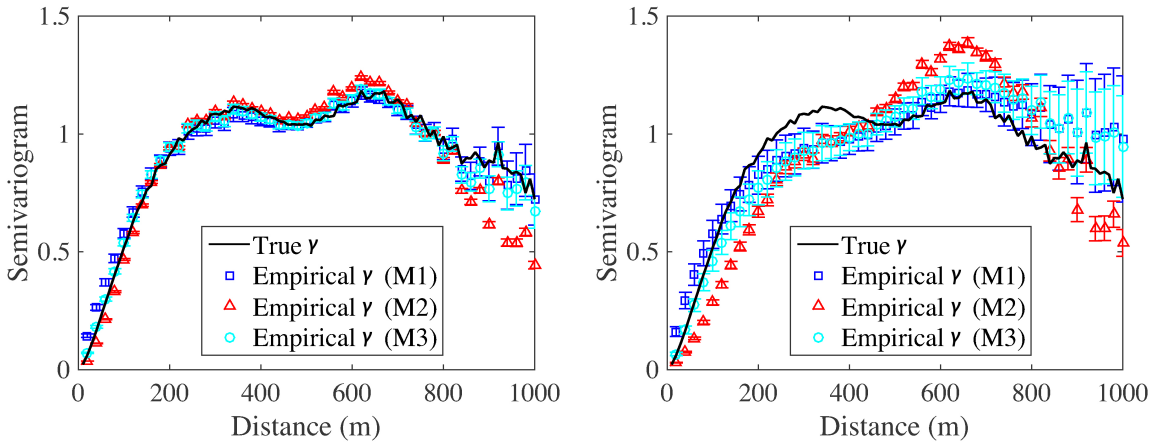
Figure 17 plots the histograms of the true and simulated LPI fields. All of the random field-based liquefaction models perform well for investigation plan #1 as the histogram fitting curves of M1, M2 and M3 are close to the true LPI histogram. The prediction accuracy decreases with the reduction in sample size, as indicated by Figure 17(b).



(a) Plan #1 (15 × 15) (b) Plan #2 (6 × 6)

Figure 17. Histograms of the true and the simulated LPI fields.

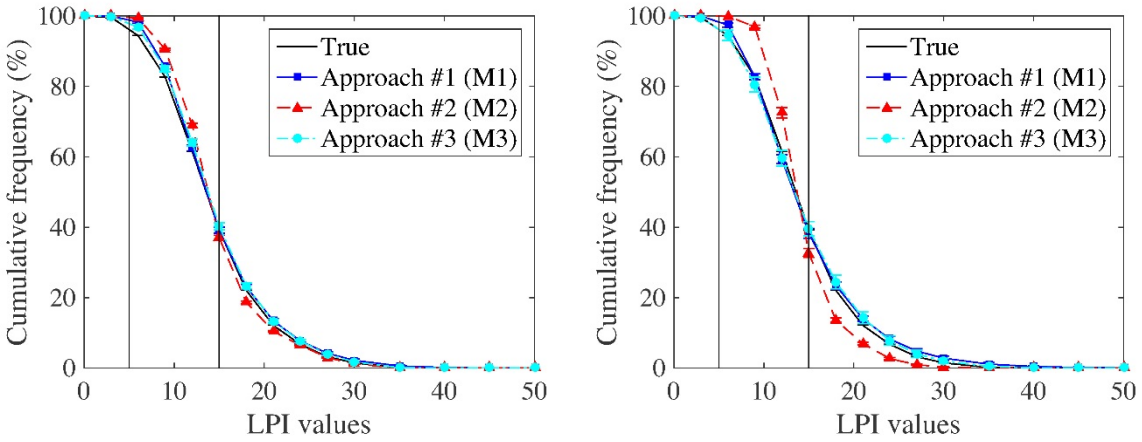
The semivariograms for simulations with investigation plan #1 and plan #2 are shown in Figure 18(a) and (b), respectively. The blue squares, red triangles and cyan circles represent the mean values of the calculated empirical semivariograms by M1, M2 and M3, respectively. The error bars indicate ±one standard deviation from the mean. It shows that the semivariograms of M1, M2 and M3 are very close to the true empirical semivariograms using sufficient samples (investigation plan #1). The variability increased when the distance of semivariogram is greater than 800 m as evidenced by longer error bars. However, the use of insufficient samples (plan #2) yields significant differences between the results of the three models and true empirical semivariograms.



(a) Plan #1 (15 × 15) (b) Plan #2 (6 × 6)

Figure 18. Semivariograms of the true and simulated LPI fields.

The performances of random field-based liquefaction models throughout the studied site are next analyzed with the cumulative frequencies shown in Figure 19. From Figure 19, it can be seen that the cumulative frequencies of M1 and M3 are very close to the true ones for both investigation plans. The model performance of M2 is worse than M1 and M3, especially under the insufficient test samples, as shown in Figure 19(b). With severity class of liquefaction defined in Table 1, it is possible to estimate the percentage of the studied site that may experience a particular level of liquefaction damage. For instance, from Figure 19(a), 96% of the studied site may experience a moderate to high liquefaction (LPI > 5) and 37% may experience a very high liquefaction (LPI > 15).

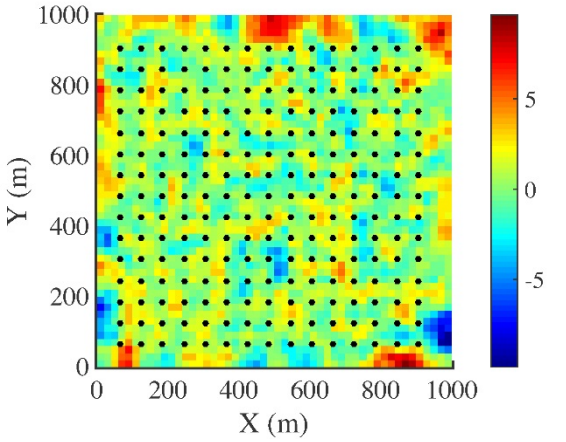


(a) Plan #1 (15 × 15) (b) Plan #2 (6 × 6)

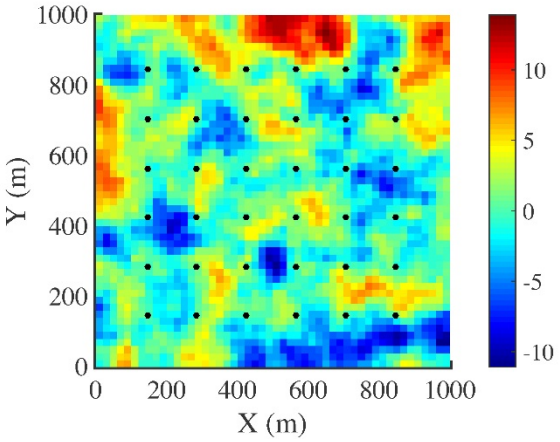
Figure 19. Cumulative frequency plots of the true and simulated LPI fields

The contours in Figure 20 are the differences between the simulated LPI values of M1, M2 and M3 and the true LPI values. The red color represents an overestimation, blue color represents an underestimation and green color represents an unbiased prediction. Observations of the contours clearly reveal that for investigation plan #1, most of the areas are within the unbiased or little bias region, indicating good model performances of the three random field

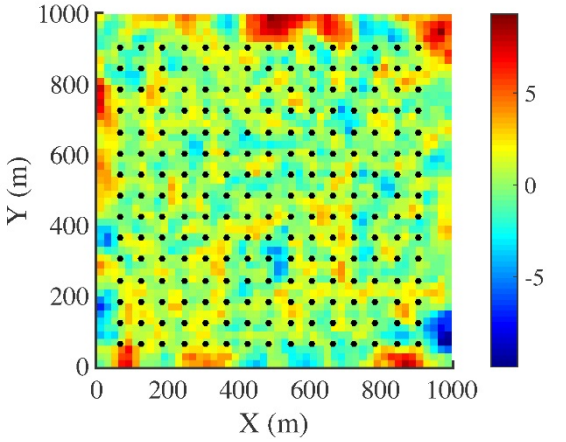
models. Over- and underestimations happen mostly around the edges of the field due to a lack of sampling data. Again, the reduction of the sample size increases the bias of the prediction, as indicated by the contours corresponding to simulations with plan #2 data.



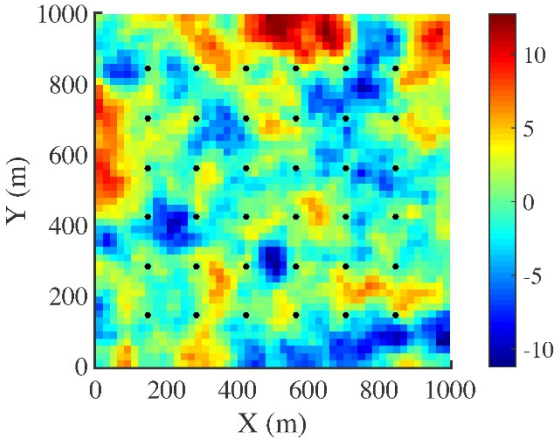
(a) M1, plan #1



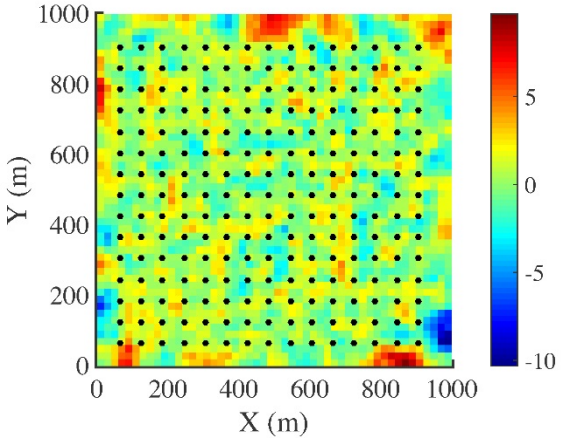
(b) M1, plan #2



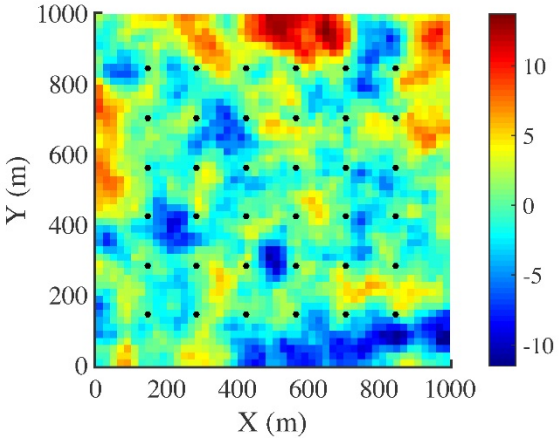
(c) M2, plan #1



(d) M2, plan #2



(e) M3, plan #1



(f) M3, plan #2

Figure 20. Contours of simulated LPI values (LPI_{sim}) minus the true LPI values (LPI_{true}) for both investigation plans

The model performances of LPI field are also quantitatively assessed with the MAPE, RMSD and bias factor. The calculated indices for M1, M2 and M3 are summarized in Table 3. All three models predict relatively accurate LPI values over the entire field when the data from plan #1 are used. The M3 outperforms M1, and the latter outperforms M2. All three approaches slightly overestimate the LPI field as the bias factors are all greater than one. When the number of test samples is insufficient (Plan #2), the model performances based on MAPE and bias factor are M3 (best), M2 (second), and M1 (worst). By RMSD, however, M2 (3.876) is slightly better than M3 (3.897) and better than M1 (3.965). It indicates that the local soil property approach (M2 and M3) is superior to the averaged index approach (M1) in predicting the liquefaction potential field when the sampling data is insufficient.

Table 3. The criteria index for LPI random fields

Index	Approach 1 (M1)		Approach 2 (M2)		Approach 3 (M3)	
	Plan #1	Plan #2	Plan #1	Plan #2	Plan #1	Plan #2
MAPE	0.146	0.300	0.150	0.293	0.130	0.285
RMSD	2.030	3.965	2.047	3.876	1.878	3.897
Bias factor	1.085	1.151	1.088	1.138	1.070	1.120

The computational efficiency of a model is also of concern when evaluating the model performance. The computational time required for obtaining the 1000 LPI random fields based on investigation plan #2 are 4.5 mins, 1090.6 mins, and 5237.4 mins for M1, M2, and M3, respectively. These numbers are meaningful only on a relative basis, as they depend on the computer used in the computation. The averaged index approach (M1) dominates in terms of computational efficiency and would be the clear choice when computational cost is of major concern.

In this work, data inference and model verification are carried out based on a synthetic digital soil field. The synthetic field affords us extremely detailed information on soil properties and a benchmark field for liquefaction potential. The focus of this work is on understanding and verifying different random field-based approaches for liquefaction potential mapping, and use of the synthetic digital field in this fundamental study has distinctive advantages over any real-world site investigation data.

On the other hand, it is important to note the assumptions and limitations of the synthetic field and the associated model verification process when drawing conclusions from the analysis. For instance, in preparing the synthetic field and in generating random field models, stationarity of the random field is assumed. Soil properties are assumed to be isotropic on a horizontal plane and anisotropic on a vertical plane. In reality, non-stationary variations of soil properties are quite common. In addition, only evenly spaced virtual field investigation plans are considered in this study, which simplifies the inference of random field model parameters. In real-world field investigations, unevenly spaced and/or clustered sampling locations are common in engineering practice. Further study to consider the effect of unevenly spaced and/or clustered sampling plans

on the data inference and model verification processes and outcomes for random field-based liquefaction hazard mapping is warranted.

4 PROJECT DATA

A data repository has been created on Github that holds data and scripts generated from this project. The repository contains data used to generate figures and liquefaction maps presented in this report. The repository can be accessed and downloaded using the web URL: https://github.com/qschen/multiscale-liquefaction

5 SUMMARY

In this one-year project, the project team developed a multiscale random field model-based methodology to map liquefaction potentials across a region. The method accounts for the spatial variability of soil parameters or the estimated liquefaction potential across different length scales. Both geotechnical data and surficial geology data-based liquefaction potential estimations are integrated in the mapping process through a conditional simulation algorithm. To verify and compare the effectiveness of different random field model-based mapping methods, a synthetic digital soil field is developed and a benchmark liquefaction potential field is used to quantify the performance of different mapping approaches. Main findings of the project are summarized as:

From the geotechnical and geology-data based mapping study, it is found that

- The proposed conditional simulation algorithm and the adoption of the Markov-Bayes coefficient can effectively account for both geotechnical data and surficial geologic data-based mapping study.
- The influence of including surficial geological data is significant. For the study site of Alameda County, without geological constraints and with only geotechnical data, the generated liquefaction hazard map incorrectly predicts high liquefaction hazard in the bedrock geologic unit.
- As the weight of the geological data increases, the geological boundaries become more distinguishable in the generated liquefaction hazard map.
- With an appropriately calibrated Markov-Bayes coefficient (B=0.73 for the study site), the accuracy of the liquefaction potential map is improved when validated against prior knowledge and liquefaction observations of the region.

From the synthetic digital soil field-based verification study, it is found that

- All three random field models examined can capture closely the statistical distribution and spatial structure of the true $(qc1N)_{cs}$ and LPI fields, provided that the amount of field test data for model parameter inference is sufficient. The model performances deteriorate with the reduction of test samples as expected.
- All random field models are found to overestimate slightly liquefaction potentials over the studied area, compared to the benchmark liquefaction potential fields.
- When there is a sufficient amount of field data for model parameter inference, the 3D local soil property approach (M3) slightly outperforms the averaged index approach (M1) and the

2D local soil property approach (M2) in terms of the accuracy in predicting the liquefaction potentials, while M1 is significantly more efficient than M2 and M3.

- When there are sufficient field test data to infer model parameters, it is recommended that the averaged index approach (M1) be used for liquefaction mapping considering a tradeoff between efficiency (in terms of computational effort) and accuracy. On the other hand, under the scenario of insufficient data, the 3D local soil property approach (M3) is recommended for its highest accuracy among the three models examined.

It should be noted that the above conclusions on model verification were reached using a synthetic digital field with the assumptions of stationarity of the random field and evenly-spaced virtual field investigation plans. Thus, these conclusions should be viewed with caution and further study to quantify the effect of these assumptions is warranted.

Ongoing research focuses on the uncertain future earthquake events and is integrating the model (e.g., Zhang et al., 2018) to calculate the joint distribution of the peak ground acceleration at the surface ($a_{\max}$) and the moment magnitude of the earthquake (M_{w}).

APPENDIX

A.1 Robertson & Wride 1998 CPT model

CSR represents the expected cyclic loading on a soil. It is a function of both the properties of the expected earthquake hazard and the properties of the soil, which can be evaluated as (Youd et al., 2001)

$$CSR = \left(\frac{\tau_{av}}{\sigma_{vo}'} \right) = 0.65 \left(\frac{a_{\max}}{g} \right) \left(\frac{\sigma_{vo}}{\sigma_{vo}'} \right) r_d \left(\frac{1}{MSF} \right) \left(\frac{1}{K_d} \right) \tag{20}$$

where $a_{\max}$ is the peak horizontal acceleration at the ground surface generated by the earthquake; g is the acceleration of gravity; σ_{vo} and σ'_{vo} are total and overburden stresses, respectively.

The variable r_d is the stress reduction coefficient, defined below (z is the depth in meters):

$$r_d = \frac{1 - 0.4113z^{0.5} + 0.04052z + 0.001753z^{1.5}}{1 - 0.4177z^{0.5} + 0.05729z - 0.006205z^{1.5} + 0.001210z^2} \tag{21}$$

MSF refers to the magnitude scaling factor. The MSF value is used to normalize the intensity of the chosen design earthquake to a value that can be used with the CSR equation, which was developed for M_w =7.5 events. The MSF can be calculated as follows:

$$MSF = \frac{10^{2.24}}{M_w^{2.56}} \tag{22}$$

Finally, K_σ is a correction factor developed by Seed (1981) to account for the nonlinear effect of overburden pressure:

$$K_{\sigma} = \left(\frac{\sigma_{vo}'}{P_a} \right)^{(f-1)} \tag{23}$$

Where P_a is the atmospheric pressure, and f is an exponent used to reflect site conditions. Youd et al. (2001) recommend that the exponent f be given a value of 0.7 to 0.8 for relative densities between 40 and 60%, and a value of 0.6 to 0.7 for relative densities between 60 and 80%.

CRR is used to quantify the liquefaction resistance of a soil. The CPT-based CRR calculation used in this methodology was developed by Robertson and Wride (1998), and can be approximated with the following equation:

$$CRR = \begin{cases} 0.833\left[\dfrac{(q_{c1N})_{cs}}{1000} \right] + 0.05; & if\ (q_{c1N})_{cs} < 50 \\ 93\left[\dfrac{(q_{c1N})_{cs}}{1000} \right]^3 + 0.08; & if\ 50 \le (q_{c1N})_{cs} < 160 \end{cases} \tag{24}$$

Where $(q_{c1N})_{cs}$ is the clean-sand equivalent normalized cone penetration tip resistance. The calculation of $(q_{c1N})_{cs}$ recommended in Youd et al. (2001) is used

$$(q_{c1N})_{cs} = K_c (q_{c1N}) \tag{25}$$

where the conversion factor K_c is calculated from the soil behavior type index I_c as

$$K_c = \begin{cases} 1 & for\ I_c \le 1.64 \\ -0.403 I_c^4 + 5.581 I_c^3 - 21.63 I_c^2 + 33.75 I_c - 17.88 & for\ I_c > 1.64 \end{cases} \tag{26}$$

and q_{c1N} is the normalized cone penetration resistance

$$q_{c1N} = \left(\frac{q_c - \sigma_{vo}}{P_a} \right) \left(\frac{P_a}{\sigma_{vo}'} \right)^n \tag{27}$$

where P_a is the atmospheric pressure; q_c is the measured cone penetration resistance; σ_{vo} and σ'_{vo} are the total and effective vertical overburden stresses, respectively. The stress exponent n is estimated as (Robertson 2009):

$$n = 0.381 I_c + 0.05 \left(\frac{\sigma_{vo}'}{P_a} \right) - 0.15 \quad \text{where } n \le 1 \tag{28}$$

in which I_c is the soil type index defined by Robertson and Wride (1998).

A.2 The liquefaction potential index model by Iwasaki

Liquefaction Potential Index (LPI) is an index that was developed by Iwasaki et al. (1978, 1982) to provide a simple evaluation of the properties of the top 20 meters of soil and the

potential that, given a seismic event, liquefaction induced site effects might occur. The index utilizes soil layer and FS information as its inputs, as follows

$$LPI = \int_0^{20} [\omega(z) \cdot F_L] dz \tag{29}$$

where z is the soil depth in meters, F_L is related to the factor of safety against liquefaction (FS) as (Sonmez, 2003)

$$F_L = \begin{cases} 0, \text{ if } FS \geq 1.2 \\ 1 - FS, \text{ if } FS \leq 0.95 \\ 2 \times 10^6 e^{-18.427 FS}, \text{ if } 0.95 \leq FS < 1.2 \end{cases} \tag{30}$$

and $\omega(z)$ is a depth weighting factor, defined as follows

$$\omega(z) = 10 - 0.5z \tag{31}$$

This weighting factor is included to account for the decreasing effect of liquefaction that occurs with depth. In other words, even if liquefaction does occur in a deep soil, as the depth of the liquefied layer below the ground surface increases, the chance of surface manifestation of that liquefaction decreases. Therefore, the value of $\omega(z)$ decreases with increasing depth to a value of zero when the depth reaches 20 meters.

The LPI can be used to classify the severity of liquefaction, where higher values of LPI correspond to greater degrees of severity as summarized in Table 4.

Table 4. Classification of the Liquefaction Potential Index (LPI) (Sonmez, 2003)

LPI Range	Severity class
LPI = 0	I: Non-liquefiable
0 < LPI ≤ 2	II: Low
2 < LPI ≤ 5	III: Moderate
5 < LPI ≤ 15	IV: High
LPI > 15	V: Very high

BIBLIOGRAPHY OF ALL PUBLICATIONS

Wang, C. (2017). Multiscale regional liquefaction assessment and mapping. Ph.D. Dissertation, Clemson University.

Juang, C. H., Shen, M., Wang, C., and Chen, Q. (2018). Random field-based regional liquefaction hazard mapping—data inference and model verification using a synthetic digital soil field. *Bulletin of Engineering Geology and the Environment*, **77**(3):1273-1286, doi:10.1007/s10064-017-1071-y.

Wang, C., and Chen, Q. (2018). A hybrid geotechnical and geological data-based framework for multiscale regional liquefaction hazard mapping. *Géotechnique*, *68*(7), 614-625, doi: 10.1680/jgeot.17.P.074.

Chen, Q., Wang, C., Lai, Z., and Juang, C.H. (2018). Integration of heterogeneous data for multiscale regional liquefaction settlement mapping. *Proceedings of the 5th Geotechnical Earthquake Engineering and Soil Dynamics Conference,* Austin, TX.

Wang, C., Chen, Q., Juang, C. H., and Liu, F. (2018). Multiscale Regional Liquefaction Hazard Mapping Accounting for Heterogeneous Data Sources. In *GeoShanghai International Conference*, pp. 44-52, doi:10.1007/978-981-13-0131-5_5, Springer, Singapore.

Shen, M., Chen, Q., Juang, C. H., Gong, W., and Tan, X. (2018). Bi-objective Optimization of Site Investigation Program for Liquefaction Hazard Mapping. In *GeoShanghai International Conference*, pp. 86-93, doi:10.1007/978-981-13-0131-5_10. Springer, Singapore.

REFERENCES

Armstrong, J. S., and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, *8*(1), 69-80.

Baker, J. W., Seifried, A., Andrade, J. E., and Chen, Q. (2011). Characterization of random fields at multiple scales: an efficient conditional simulation procedure and applications in geomechanics. *11th International Conference on Applications of Statistics and Probability in Soil and Structural Engineering (ICASP11)*, Zurich, Switzerland. 7p.

Baise, L.G., Higgins, R.B., and Brankman, C.M. (2006). Liquefaction hazard mapping—statistical and spatial characterization of susceptible units. *Journal of Geotechnical and Geoenvironmental Engineering*, 132(6), 705-715.

Chen, Q., Seifried, A., Andrade, J. E., and Baker, J. W. (2012). Characterization of random fields and their impact on the mechanics of geosystems at multiple scales. *International Journal for Numerical and Analytical Methods in Geomechanics*, *36*(2), 140-165.

Chen, Q., Wang, C., and Hsein Juang, C. (2016a). CPT-based evaluation of liquefaction potential accounting for soil spatial variability at multiple scales. *Journal of Geotechnical and Geoenvironmental Engineering*, *142*(2), 04015077.

Chen, Q., Wang, C., and Juang, C. H. (2016b). Probabilistic and spatial assessment of liquefaction-induced settlements through multiscale random field models. *Engineering Geology*, *211*, 135-149.

Chen, Q., Wang, C., Lai, Z., and Juang, C.H. (2018). Integration of heterogeneous data for multiscale regional liquefaction settlement mapping. *Proceedings of the 5th Geotechnical Earthquake Engineering and Soil Dynamics Conference,* Austin, TX.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, *17*(5), 563-586.

Goovaerts, P. (1997). Geostatistics for natural resources evaluation. New York, NY, USA: Oxford University Press.

Heidari, T. and Andrus, R. D. (2010). Mapping liquefaction potential of aged soil deposits in Mount Pleasant, South Carolina. *Engineering Geology*, 112, 1–12.

Helley, E. J. and Graymer, R. W. (1997). Quaternary geology of Alameda County, and parts of Contra Costa, Santa Clara, San Mateo, San Francisco, Stanislaus, and San Joaquin counties, California: A digital database, Technical Report USGS Open-File Report 97-97. Reston, VA, USA: United States Geological Survey.

Holzer, T. L., Bennett, M. J., Noce, T. E., Padovani, A. C. and Tinsley III, J. C. (2006). Liquefaction hazard mapping with LPI in the greater Oakland, California, area. *Earthquake Spectra*, 22 , 693–708.

Iwasaki, T., Tatsuoka, F., Tokida, K. I., and Yasuda, S. (1978). A practical method for assessing soil liquefaction potential based on case studies at various sites in Japan. *Proceeding of the 2nd International Conference on Microzonation for Safer Construction-Research and Application*, Vol. II, San Francisco, 885–896.

Iwasaki, T., Tokida, K., Tatsuoka, F.,Watanabe, S., Yasuda, S., and Sato, H. (1982). Microzonation for soil liquefaction potential using simplified methods. *Proceeding of the 3rd International Conference on Microzonation*, Vol. 3, Seattle,1319–1330.

Hsein Juang, C., Luo, Z., Atamturktur, S., and Huang, H. (2012). Bayesian updating of soil parameters for braced excavations using field observations. *Journal of Geotechnical and Geoenvironmental Engineering*, *139*(3), 395-406.

Juang, C. H., Shen, M., Wang, C., and Chen, Q. (2018). Random field-based regional liquefaction hazard mapping—data inference and model verification using a synthetic digital soil field. *Bulletin of Engineering Geology and the Environment*, **77**(3):1273-1286, doi:10.1007/s10064-017-1071-y.

Kayen, R. E., Mitchell, J. K., Seed, R. B. & Nishio, S. (1998). Soil liquefaction in the east bay during the earthquake. In The Loma Prieta, California earthquake – liquefaction (ed. T. L. Holzer), Professional Paper No. 1551-B, pp. 61–86. Denver, CO, USA: United States Geological Survey.

Knudsen, K. L., Sowers, J. M., Witter, R. C., Wentworth, C. M., and Helley, E. J. (2000). Description of mapping of quaternary deposits and liquefaction susceptibility, nine-county San Francisco Bay Region, California. *Rep. No. United States Geologic Survey Open-File Report 00*, *444*.

Kung, G. T., Juang, C. H., Hsiao, E. C., and Hashash, Y. M. (2007). Simplified model for wall deflection and ground-surface settlement caused by braced excavation in clays. *Journal of Geotechnical and Geoenvironmental Engineering*, *133*(6), 731-747.

Lenz, J.A., and Baise, L.G. (2007). Spatial variability of liquefaction potential in regional mapping using CPT and SPT data. *Soil Dynamics and Earthquake Engineering*, 27, 690-702.

Liu, C. N., and Chen, C. H. (2006). Mapping liquefaction potential considering spatial correlations of CPT measurements. *Journal of Geotechnical and Geoenvironmental Engineering*, *132*(9), 1178-1187.

Liu, W., Chen, Q., Wang, C., Juang, C. H., and Chen, G. (2017). Spatially correlated multiscale Vs30 mapping and a case study of the Suzhou site. *Engineering Geology*, *220*, 110-122.

Moysey, S., Caers, J., Knight, R., and Allen-King, R. M. (2003). Stochastic estimation of facies using ground penetrating radar data. *Stochastic Environmental Research and Risk Assessment*, *17*(5), 306-318.

Prasomphan, S., and Mase, S. (2013). Generating prediction map for geostatistical data based on an adaptive neural network using only nearest neighbors. *International Journal of Machine Learning and Computing*, *3*(1), 98.

Robertson, P. K., and Wride, C. E. (1998). Evaluating cyclic liquefaction potential using the cone penetration test. *Canadian Geotechnical Journal*, *35*(3), 442-459.

Robertson PK (2009) Performance based earthquake design using the CPT. Proc. IS-Tokyo, pp 3–20.

Shen, M., Chen, Q., Juang, C. H., Gong, W., and Tan, X. (2018). Bi-objective Optimization of Site Investigation Program for Liquefaction Hazard Mapping. In *GeoShanghai International Conference*, pp. 86-93, doi:10.1007/978-981-13-0131-5_10. Springer, Singapore.

Sonmez, H. (2003). Modification of the liquefaction potential index and liquefaction susceptibility mapping for a liquefaction-prone area (Inegol, Turkey). *Environmental Geology*, *44*(7), 862-871.

Vivek, B., and Raychowdhury, P. (2014). Probabilistic and spatial liquefaction analysis using CPT data: a case study for Alameda County site. *Natural Hazards*, *71*(3), 1715-1732.

Wang, C., Chen, Q., Shen, M., & Juang, C. H. (2017). On the spatial variability of CPT-based geotechnical parameters for regional liquefaction evaluation. *Soil Dynamics and Earthquake Engineering*, *95*, 153-166.

Wang, C., and Chen, Q. (2018). A hybrid geotechnical and geological data-based framework for multiscale regional liquefaction hazard mapping. *Géotechnique*, *68*(7), 614-625, doi: 10.1680/jgeot.17.P.074.

Wang, C., Chen, Q., Juang, C. H., and Liu, F. (2018). Multiscale Regional Liquefaction Hazard Mapping Accounting for Heterogeneous Data Sources. In *GeoShanghai International Conference*, pp. 44-52, doi:10.1007/978-981-13-0131-5_5, Springer, Singapore.

Webster, R., and Oliver, M. A. (1992). Sample adequately to estimate variograms of soil properties. *Journal of Soil Science*, *43*(1), 177-192.

Witter, R. C., Knudsen, K. L., Sowers, J. M., Wentworth, C. M., Koehler, R. D., Randolph, C. E., and Gans, K. D. (2006). *Maps of Quaternary deposits and liquefaction susceptibility in the central San Francisco Bay region, California*, U.S. Geological Survey Open-File Report 2006-1037.

Youd, T. L., and Perkins, D. M. (1978). Mapping liquefaction-induced ground failure potential. *Journal of Geotechnical and Engineering Division*, 104(4), 433-446.

Youd, T. L., Idriss, I. M., Andrus, R. D., Arango, I., Castro, G., Christian, J. T., Dobry, R., Finn, W. L., Harder, L. F. Jr, Hynes, M. E., Ishihara, K., Koester, J. P., Liao, S. S. C., Marcuson, W. F. III, Martin, G. R., Mitchell, J. K., Moriwaki, Y., Power, M. S., Robertson, P. K., Seed, R. B. and Stokoe, K. H. II (2001). Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. *Journal Geotechnical Geoenvironmental Engineering*, 127, No. 10, 817–833.

Zhang, J., Chen, F. Y., Juang, C. H., and Chen, Q. (2018). Developing joint distribution of amax and Mw of seismic loading for performance-based assessment of liquefaction induced structural damage. *Engineering Geology*, *232*, 1-11.

Zhu, J., Daley, D., Baise, L. G., Thompson, E. M., Wald, D. J., & Knudsen, K. L. (2015). A geospatial liquefaction model for rapid response and loss estimation. *Earthquake Spectra*, *31*(3), 1813-1837.

Zhu, J., Baise, L. G., & Thompson, E. M. (2017). An updated geospatial liquefaction model for global application. *Bulletin of the Seismological Society of America*, *107*(3), 1365-1385.